Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Conditioned Fear as a Function of CS-UCS and Probe Stimulus Intervals: L. E. ROSS 265
The Perception of All Patterns Produced by a Seven-Line Matrix: E. T. KLEMMER 274
Paired-Associate Learning as a Function of Percentage of Occurrence of Response Members and Other Factors: H. C. WILCOXON, W. R. WILSON, AND D. A. WISE 283
The Effect of CS Presence during Delay of Reward on the Speed of an Instrumental Response: M. RIEBER 290
Induced Muscular Tension, Incentive, and Blink Rate in a Verbal Learning Task: J. B. SIDOWSKI AND C. NUTHMANN 295
Denotative Meaning Established by Classical Conditioning: A. W. STAATS, C. K. STAATS, AND W. G. HEARD 300
Resistance to Extinction as a Function of the Fixed Ratio: J. J. BOREN 304
Differential Relation of Latency and Response Vigor to Stimulus Similarity in Brightness Discrimination: A. CASTANEDA AND L. WORELL 309
Paired-Associate Learning when the Same Items Occur as Stimuli and Responses: R. K. YOUNG 315
The Effect of Continuous, Partial, and Varied Magnitude Reinforcement on Acquisition and Extinction: H. G. YAMAGUCHI 319
Acquisition and Transfer in Pattern-vs.-Component Discrimination Learning: W. K. ESTES AND B. L. HOPKINS 322
Concept Identification as a Function of Intra- and Interdimensional Variability: W. F. BATTIG AND L. E. BOURNE, JR. 329
An Alternative Criterion for the Elimination of "Voluntary" Responses in Eyelid Conditioning: T. F. HARTMAN AND L. E. ROSS 334
Secondary Reinforcement Based on Stimulus-Change Primary Reinforcement: C. L. ROBERTS, K. E. LEBOW, AND R. M. YODER 339
The Role of Overlearning Trials in Determining Resistance to Extinction: N. R. MURILLO AND E. J. CAPALDI 345
Supplementary Report: Resistance to Extinction of a Verbal Response as a Function of the Number of Acquisition Trials: M. G. SENKO, R. A. CHAMP, AND E. J. CAPALDI 350
Supplementary Report: Spontaneous Recovery of Problem Solving Set: W. N. RUNQUIST AND B. SEXTON 351

American Psychological Association

Vol. 61, No. 4 April 1961

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 61, NO. 4 APRIL 1961

CONDITIONED FEAR AS A FUNCTION OF CS–UCS AND PROBE STIMULUS INTERVALS[1]

LEONARD E. ROSS[2]

State University of Iowa

It is generally assumed that when a neutral stimulus (CS) is paired with a noxious stimulus (UCS) the emotional (pain) response aroused by the UCS becomes conditioned to the CS. Thus the presentation of the CS is hypothesized to elicit an anticipatory emotional response (fear) which has motivational concomitants and which can, when reduced in strength, reinforce the immediately preceding response. A number of studies (e.g., Brown & Jacobs, 1949; Kalish, 1954; Miller, 1948) have supported this formulation.

Since fear is assumed to have the characteristics of a response it should be a function of those factors that affect the acquisition, strength, and maintenance of other responses (Miller, 1951). One of the variables known to be important in the conditioning of skeletal responses is the interval between the onsets of the CS and UCS, the experimental evidence clearly indicating the optimum interval to be approximately ½ sec. (McAllister, 1953; Reynolds, 1945; Spooner & Kellogg, 1947; Wolfle, 1932). In contrast to this well established relationship, the results of studies investigating the role of this variable in fear conditioning have been quite inconsistent. In part this finding is probably due to the wide variety of experimental conditions used. Thus these investigations have involved both conditioned responses (e.g., the GSR) under the control of the autonomic nervous system and instrumental responses, involving striated muscles, motivated by conditioned fear.

The evidence in regard to the conditioning of autonomic responses is conflicting. The optimum interval for GSR conditioning has been reported to be less than 1 sec. with no conditioning beyond this period (Moeller, 1954; White & Schlosberg, 1952). On the other hand two experiments have demonstrated heart-rate conditioning with a 6-sec. CS-UCS interval, while a third obtained conditioning with both 5- and 20-sec. intervals (Church & Black, 1958; Notterman, Schoenfield, & Bersh, 1952; Zeaman, Deane, & Wegner, 1954). In the conditioning of another autonomically controlled response, the dilation of the human pupil, a 1.5-sec. interval has been found superior to .125 or .5 sec. with no conditioning reported at 2.5 sec. (Gerall & Woodward, 1958).

In the case of the GSR studies Brown has suggested (Moeller, 1954) that the response conditioned was not the GSR but a

[1] This study is based on a dissertation submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the PhD degree. The writer is indebted to Kenneth W. Spence for advice and assistance throughout the course of the investigation which is part of a project concerned with the influence of motivation on performance in learning under Contract N9 onr-93802 Project NR 154-107 between the State University of Iowa and the Office of Naval Research.

[2] Now at the University of Wisconsin.

simple skeletal response which resulted in a GSR, the interval-conditioning relationship thus resembling that obtained with skeletal responses. While Smith (1954) further proposed that all autonomic conditioning is an artifact of skeletal responding, the heart rate and pupillary conditioning obtained with intervals longer than those found effective in skeletal conditioning makes this position questionable, although there is no evidence that the relationships found are entirely independent of skeletal activity.

A number of studies have utilized instrumental responses as fear indicants in investigating the CS-UCS interval. Murfin (1953) found a 5-sec. interval superior to .5- or 500-sec. intervals in a situation in which the latency of hurdle jumping during a test period served as the measure of fear conditioning. In another experiment using an instrumental response (Libby, 1951), the interfering or inhibitory effects of conditioned fear on bar pressing were found to increase as the CS-UCS interval increased from 0 to 10 sec. and then to decrease slightly as the interval increased to 30 sec. Other studies have investigated the CS-UCS interval variable in instrumental escape-avoidance situations.

The use of an instrumental response as an indicant of conditioned fear appears to suffer from several complicating factors (see Brown, Kalish, & Farber, 1951) in that instrumental response measures presumedly reflect not only the effects of conditioned fear but also factors connected with the acquisition and strength of the instrumental response itself. As a consequence this type of study, while providing valuable information concerning the relationships between the interval variable and aspects of the instrumental response, has limited relevance for the laws relating the interval variable to the conditioned fear which presumably, in part, elicits and energizes the instrumental response.

One procedure that has not been employed to investigate systematically the CS-UCS interval is the probe-stimulus technique, which is based on the intensifying or motivating properties of fear on skeletal reflex responses. Typically with this method a probe stimulus is introduced in place of the UCS on test trials during conditioning, the index of the strength of fear being the facilitating effect of the CS upon the amplitude of a reflex response to the probe stimulus. Studies using the probe-stimulus method with animals have demonstrated the facilitating effects of conditioned fear on a startle response with 3- and 20-sec. CS-UCS intervals (Brown, Kalish, & Farber, 1951; Meryman, 1952), while experiments with human *Ss* have shown facilitation of both unconditioned GSRs elicited by auditory stimuli (Meryman, 1953) and reflex blinks to air puffs (Spence & Runquist, 1958), 20- and .5-sec. CS-UCS intervals being used, respectively.

Although these experiments using the probe stimulus procedure indicate that fear conditioning is possible over a wide range of interval values, they do not allow a comparison of the relative effectiveness of different CS-UCS intervals. Such information would appear to be of special value since the probe stimulus technique avoids some of the complications inherent in procedures employing instrumental responses reinforced by fear reduction. Another advantage of the probe stimulus method is the possibility of obtaining data concerning the growth of the fear response following the CS onset by introducing the probe stimulus at different CS-probe intervals.

The present experiment was designed to investigate the nature of the laws relating the CS-UCS interval to the strength of the conditioned fear response, as well as to provide information concerning the strength of fear at different intervals following the onset of the CS. The probe stimulus technique was employed with human *Ss*, the amplitude of the reflex blink to a probe stimulus (air puff) serving as the index of conditioned fear.

METHOD

Subjects.—The *Ss* were 171 men from an introductory course in psychology. Twenty-five *Ss* who did not respond with adequate amplitude to the air puff probe during the puff adaptation period were discarded, as were 17 *Ss* who did not adapt a sufficient amount (see Procedure). Six additional *Ss* were eliminated due to a high blink rate that precluded accurate measurement of the blink to the air puff; while three other *Ss* were discarded due to apparatus failure or *E* error. The remaining 120 *Ss* were assigned to five CS-UCS interval groups with an attempt being made to equate the groups in mean amplitude of reflex blink made on the last five puff adaptation trials. Upon assignment to a CS-UCS interval group each *S* was further assigned, at random, to one of four probe-order subgroups. Each

CS-UCS group was thus comprised of 24 *Ss* with six in each subgroup.

Apparatus.—The *Ss* were seated in an adjustable dental chair located in a semi-soundproof room which was illuminated by a shielded 7½-w. bulb. The recording apparatus and stimulus controls were housed in a separate room. The apparatus for recording the eyelid reflex was a potentiometer-polygraph system described elsewhere (Spence & Taylor, 1951).

The CS was an increase in brightness of a 2½-in. dia. circular milk glass disk located 51 in. in front of and slightly above the eye level of *S*. Duration of the CS was 250 msec. The shock UCS, provided by the discharge of a 4-mfd. condenser, was administered through electrodes attached to the index and middle fingers of the right hand. Electrode paste was used to insure contact over the entire surface of the electrodes. The probe stimulus was a 50-msec., 1.25-psi air puff delivered to the left temple through a 5-mm. dia. orifice. The air-puff nozzle was placed approximately 1 in. from the corner of the eye, ¼ in. from the skin, on a horizontal plane with the eye.

Experimental design.—The CS-UCS and CS-probe intervals used were .5, 2, 5, and 10 sec. In addition a control group received the UCS 1 sec. prior to the CS. The CS-UCS interval was varied between groups of *Ss* while each *S*, including those in the backward conditioning group, received all four of the CS-probe intervals. Thirty-four conditioning trials (C) and 16 test trials (T) were given in the following order: CCTCTCCTTC TCTCCCTTCC TCTCCCTCTC CTCCTCTCCT.

Within each CS-UCS group the order of presentation of the CS-probe intervals was counterbalanced for each block of 10 trials by giving each of the subgroups one of the four possible probe interval orders. Thus on each block of 10 trials four subgroups, one from each CS-UCS group, received one order, four another order, etc. Considering a single CS-UCS group, each of the four subgroups of six *Ss* received a different order of CS-probe intervals on the four test trials of each 10-trial block.

From one 10-trial block to another the particular order received by a subgroup changed, the orders themselves being counterbalanced within each CS-UCS group over the four blocks of trials. Each subgroup therefore experienced all four orders used, but no two subgroups within the same CS-UCS group had the same probe order on the same 10-trial block.

Procedure.—Following the reading of the instructions the shock level was adjusted for each *S*. The initial presentation was at 50 v. followed by step increases to 70, 90, or 110 v. depending upon when *S* reported the shock to be definitely uncomfortable but not painful. After the final level was determined additional shocks were administered, if needed, in order to equate the number of adjustment shocks received by all *Ss* to four.

Since the index of fear used in this study was the facilitation of the eyeblink response, it was necessary to adapt this response to a low level before test trials were given. To accomplish this each *S* next received 30 puff adaptation trials at prearranged intervals of 15, 20, or 25 sec. Those *Ss* who gave five responses of less than 1 mm. in amplitude during the first 10 trials were discarded, as were *Ss* who did not adapt to a mean amplitude on Trials 26–30 that was less than 60% of their mean amplitude on adaptation Trials 1–2. In both cases *Ss* were dropped at the end of the adaptation period and did not receive conditioning or test trials.

Upon completion of the adaptation trials the conditioning and test trials were given at prearranged 30, 35, or 40 sec. intervals, the intertrial interval being defined as the interval from the UCS or probe stimulus to the onset of the following CS. For the backward conditioning group the intertrial interval was the period from the CS or probe to the following shock.

Results

Adaptation to the air puff probe.—The mean amplitude of the eyelid closures made by *Ss* in the five CS-UCS groups during puff adaptation Trials 1–2 and 26–30 is shown in Fig. 1. As may be seen the *Ss* adapted to a low level, the mean amplitude of response for each group on Trials 26–30 being less than 4 mm. Although the CS-UCS groups were matched only on the basis of response amplitude on Trials 26–30, the groups were also very similar on Trials 1–2.

Response on CS-probe trials.—Introduction of the CS-probe test trials resulted in a marked increase in response amplitude to the air puff on the first such presentation, the increase in amplitude bearing a positive relationship to the length of the

CS-probe interval. As the test trials continued this increased response amplitude was maintained or increased slightly by some groups at some probe intervals, while for other combinations of conditions the response amplitude decreased by differing amounts. These changes presumably represent the joint effects of the acquisition of conditioned fear with its attendant facilitating effects and the adaptation of the blink response to the CS-probe presentations. Thus in the present experiment the development of the facilitating effects of the conditioned fear response was revealed primarily by the maintenance of response amplitude over the series of test trials rather than by an actual increase in the amplitude of the test responses.

The data selected for analysis were those obtained on the last block of four test trials (interspersed during Trials 31–40), a point at which the conditioned fear response would be near asymptotic strength and adaptation to the CS-air puff combinations would be maximum. The mean amplitude of response made on the last four test trials by the *Ss* of

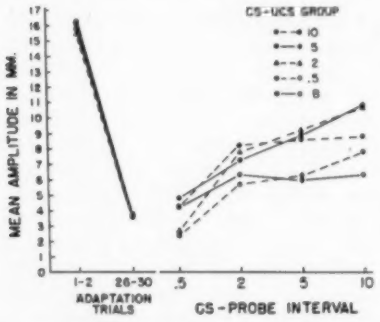

FIG. 1. Mean amplitude of response on the last test trial at each probe interval, and the mean amplitude of response on puff adaptation Trials 1–2 and 26–30, for *Ss* of each group.

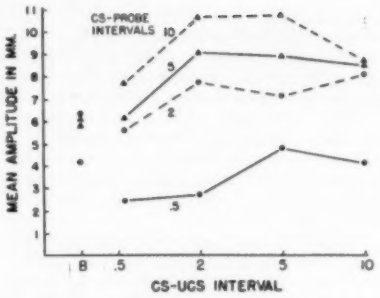

FIG. 2. Mean amplitude of response as a function of CS-UCS interval for the last test trial at each CS-probe interval.

the five CS-UCS groups is presented in Fig. 1, response amplitude being shown as a function of the CS-probe interval with the CS-UCS variable the parameter. Inspection of these curves reveals that the mean amplitude of response for the 2-, 5-, and 10-sec. groups increased as the CS-probe interval increased, the curves for these groups diverging from that of the backward group. While the .5-sec. group also shows an increase with increasing probe interval, its response level is lower than that of the backward group at both the .5- and 2-sec. probe intervals. In Fig. 2 the mean amplitude of response is shown as a function of the CS-UCS interval with the CS-probe variable the parameter.

The data were evaluated statistically by an analysis of covariance[3] involving one between factor (CS-UCS interval) and two within factors (probe interval and order), the mean amplitude of response made on puff adaptation Trials 26–30 serving as the

[3] Each time that an analysis of covariance procedure was employed in this study a test of the homogeneity of regressions of the criterion variable on the concomitant variable was carried out. In no case could the hypothesis of equal regressions be rejected at the .20 level of confidence.

concomitant variable. Since none of the interactions with order were significant (all Fs < 1) the mean squares for these interactions were pooled in the appropriate error terms. The main effects of probe interval and CS-UCS interval were significant at the .001 and .05 levels, respectively ($F = 36.96$, $df = 3/342$; and $F = 2.45$, $df = 4/114$, respectively), while the interaction of these factors was significant at only the .10 level ($F = 1.56$, $df = 12/342$).

In view of the level of statistical reliability of this CS-UCS interval by CS-probe interaction, the question arises as to whether or not an evaluation of the simple effects of these variables is justified. However, it was recognized at the beginning of the experiment that the use of four rather arbitrarily selected CS-UCS intervals in addition to a quantitatively different condition (the backward group), and the testing of these at a number of probe intervals, might well result in interactions between specific groups for some probe intervals to which the over-all analysis would be relatively insensitive. It was thus decided that evaluation of the simple effects was justified; the risks of capitalizing on large chance differences in the simple effects analyses appearing to be smaller than the dangers of misinterpretation that might result from consideration of the main effects alone.[4]

Accordingly, the CS-UCS interval variable was tested at each CS-probe interval, and the effect of the CS-probe interval factor was analyzed separately for each CS-UCS group.

The effect of CS-UCS interval.—In order to evaluate the simple effects of the CS-UCS interval factor four simple randomized analyses of covariance were performed, one at each CS-probe interval.

These analyses revealed that the CS-UCS variable did not reach the .05 level at either the .5- or the 2-sec. probe interval, but that it was significant at the 5- and 10-sec. probe intervals ($P < .05$ and $P < .005$, respectively). In the case of these latter two analyses t tests were used to compare individual pairs of CS-UCS group means. At both intervals the 2- and the 5-sec. CS-UCS interval groups were significantly superior to the .5-sec. and backward groups at the .05 level. In addition the differences between the 10-sec. and backward groups approached but did not reach statistical significance at the .05 level at these probe intervals (differences obtained and required: 2.72 and 2.77, 2.67 and 2.68, respectively).

The effect of CS-probe interval.—In order to test the simple effects of the different CS-probe intervals, a probe interval by *S*s analysis of variance was carried out for each CS-UCS interval group. In each case the error term employed was the residual within *S*s error term of the over-all analysis of covariance.

The results of these analyses showed the CS-probe interval factor to be significant at the .001 level for the .5-, 2-, 5-, and 10-sec. CS-UCS groups, while the F for the backward group failed to reach the .20 level. Accordingly, t tests were made of the differences between the probe interval means for each of the forward conditioning groups. It was found that the mean amplitude of response at the .5-sec. probe interval was less than

[4] As an alternative treatment of the data, analyses of covariance (involving the same three factors as the over-all analysis) were made comparing each forward CS-UCS group separately with the backward group. The results of these analyses (Ross, 1959) lead to essentially the same conclusions as the treatment presented in this section.

that at each of the other intervals at the .05 level or less in each case, while the amplitude at the 10-sec. probe interval was significantly superior to that at the 2-sec. interval in the case of the 2- and the 5-sec. CS-UCS groups ($P < .05$). All other differences failed to reach the .05 level.

Discussion

The effect of CS-UCS interval.—The present experiment indicates that fear can be conditioned with either a 2- or 5-sec. CS-UCS interval, but that a .5-sec. interval results in little or no conditioning. These conclusions are based on the significant superiority of the 2- and 5-sec. groups when compared to the .5-sec. and backward conditioning groups at the 5- and 10-sec. probe intervals. It would also seem fairly safe to assume that fear conditioning took place in the case of the 10-sec. CS-UCS group since this group differed from the backward condition at approximately the .05 level when tested at these same probe intervals. While the response amplitude of the .5-sec. group did increase as the probe interval increased, the differences between it and the backward group all failed to reach significance. It thus appears that little or no conditioning occurred at this CS-UCS interval.

It should be noted that in a study by Spence and Runquist (1958) which employed a technique similar to that used in this experiment a .5-sec. CS-UCS group exhibited a significantly larger mean amplitude of response to the probe stimulus at 4.5 sec. than the backward group, while the difference in the present experiment at a 5-sec. probe interval was not significant. In the Spence and Runquist study, however, a ready signal preceded the CS by several seconds on each trial, and this ready signal could have acted as a trace CS for a conditioned fear response.

The effect of CS-probe interval.—The results of the analyses of CS-probe interval differences for each CS-UCS group, together with the CS-UCS differences discussed in the previous section, indicate that the latency of the conditioned fear response is longer than .5 sec., with some evidence suggesting that the fear response is still relatively small in magnitude by 2 sec. While it appears from Fig. 1 and 2 that the mean amplitude of response for the forward conditioning groups increases with each increase in probe interval, the statistical analyses of the differences between probe intervals does not support such a conclusion. Thus, while amplitude of response was significantly less at the .5-sec. interval than at the longer intervals and there were also significant increases as the probe interval increased from 2 to 10 sec., the differences between the 5- and 10-sec. intervals were not significant and, in addition, there is reason to question the apparent increase from the .5- to the 2-sec. interval.

The possibility should be considered that the increases in amplitude of response with increasing probe interval actually represent a continuation of the gradient found on the first test trial. Thus, although the over-all effect of probe interval was not significant in the case of the backward group (which was presumably free from the facilitating effects of conditioned fear), the fact that response amplitude for this group is less at the .5-sec. interval than at any of the longer intervals raises the possibility that these increases are not chance ones, and that some portion of the increases found with the forward conditions might be due to a similar unadapted increase. Accordingly, the individual forward group comparisons were recalculated with the differences reduced by the amount of increase present in the backward group. Under these conditions the differences between the .5- and 2-sec. probe intervals were no longer significant for the .5-, 5-, and 10-sec. CS-UCS groups, while the difference between the .5- and 5-sec. probe intervals now also failed to reach significance in the case of the .5-sec. CS-UCS group. It is therefore questionable that the differences between the .5- and the 2-sec. intervals represent a facilitation of the

response at the 2-sec. interval by conditioned fear. In any case the failure to find significant differences between the forward and backward groups at this probe interval suggests that if conditioned fear is present 2 sec. following the onset of the CS, it is present to a relatively small degree.

While Church and Black (1958) have reported that the conditioned heart rate response evidences inhibition of delay, there is no indication of such a phenomenon in the present study. Neither is there any diminution of the facilitating effects of fear at periods longer than the CS-UCS interval employed since the amplitude of response for the 2- and 5-sec. CS-UCS groups did not decrease, but instead showed some increase with probe intervals longer than the conditioning interval.

The facilitating effects of conditioned fear.—The results of the present experiment provide additional support for the assumption that the pairing of a neutral and a noxious stimulus results in the establishment of a conditioned fear response to the neutral stimulus, and that this fear response can energize or facilitate a reflex response. However, in contrast to other probe studies (Brown, Kalish, & Farber, 1951; Meryman, 1952, 1953) which have found an increase in response amplitude with conditioning trials, the facilitating effects of fear in the present study, as in the Spence and Runquist experiment, were manifested primarily through the maintenance, or reduced adaptation, of the fear indicant response.

It would appear that this failure to find an increase in response strength with fear conditioning trials was due to the large amplitude of response found on the initial test trials. That is, while the puff adaptation trials were effective in reducing the amplitude of response to the puff when it was presented alone, these trials had a lesser effect upon the amplitude of response to the air puff when it was preceded by the CS. Thus when the probe trials were introduced the indicant response blink was still of large amplitude.

The fact that differences were obtained through a reduction of the amount of adaptation rather than by increases in the amplitude of the indicant response with trials does not qualify the interpretation of these differences. Nevertheless it would obviously be desirable to develop a procedure for adapting the eyeblink to CS-probe combinations before the conditioning period. Not only would this allow a study of the growth of the facilitating effects of fear without the complication of a concurrent adaptation of the indicant response, but it would also insure that the differences obtained were not reduced due to incomplete adaptation at the point of comparison. It is possible, for example, that the differences between the forward and backward groups in the present experiment might have been greater after additional probe trials. A way of accomplishing such a preconditioning adaptation might be to give CS-probe pairings during the adaptation period with CS-probe intervals long enough to avoid the conditioning of an eyelid closure to the CS. The same results could possibly be obtained by giving random presentations of both the CS and air puff prior to conditioning.

It should be noted that data obtained by Kimble and Ost (1961)[5] have been interpreted by them as demonstrating an inhibitory process under CS control which reduces the amplitude of the UCR to an air puff UCS. This inhibitory process is further believed to vary with CS-UCS interval, being maximal at .5 sec. and decreasing with longer interstimulus intervals. The question arises concerning the effect such a factor might have had on the data of the present study.

First, if such inhibition affected CS-probe trials as a function only of previous CS-probe pairings, the effect should be the same for all groups since each group received the same test trial CS-probe intervals (with order counterbalanced the same for all groups). Another pos-

[5] The author would like to thank Gregory A. Kimble for making this manuscript available prior to publication.

sibility is that fear conditioning itself might have been reduced differentially as a function of CS-UCS interval by conditioned inhibition under the control of the CS. This, however, would not affect the conclusions concerning the effect of CS-UCS interval but rather would be a kind of "explanation" of the results.

A third possibility is that fear conditioning trials might have resulted in inhibition which affected the reflex responses to the air puff. That is, the .5-sec. CS-shock groups' response amplitude on CS-air puff trials was smaller than that of *S*s who received fear conditioning trials with longer CS-shock intervals because of a greater amount of conditioned inhibition. However if this were the case the response amplitude for the .5-sec. CS-UCS group would be expected to decrease more from the first to last block of test trials, as the inhibition develops, than the backward condition group which presumedly would be free from this type of inhibition. A comparison of the response amplitudes made on the first and last blocks of trials shows that this was not the case, the response amplitude decrease of the backward conditioning group being as great as or greater at each CS-probe interval than that of the .5-sec. CS-UCS group. Further, one might expect the .5-sec. groups' decrement to remain constant with increasing CS-probe interval, or possibly be reduced with recovery from inhibition. That such is not the case is apparent from Fig. 1 where the amplitude of response for the .5-sec. group diverges from the curves for the other fear conditioning groups as a function of CS-probe interval.

Summary

An experiment was conducted to investigate the nature of the function relating the strength of conditioned fear to the CS-UCS interval, and provide information concerning the strength of fear at different intervals following the onset of the CS. The probe stimulus technique was employed with human *S*s, the amplitude of the reflex blink to an air puff probe stimulus serving as the index of conditioned fear. Groups of 24 *S*s each were conditioned with different CS-UCS intervals, each *S* receiving probe stimulus test trials at various CS-probe intervals. The CS-UCS and CS-probe intervals used were .5, 2, 5, and 10 sec. A control group was employed which received the UCS 1 sec. prior to the CS. Four blocks of 10 trials each were given, each block consisting of six fear conditioning trials and four probe test trials.

Analyses of the amplitude of response made on the last four test trials were interpreted as indicating that (*a*) fear conditioning will occur with a 2-, 5-, or 10-sec. CS-UCS interval but that little if any fear conditioning occurs with a .5-sec. CS-UCS interval, and (*b*) that the latency of the conditioned fear response is longer than .5 sec. with little fear present at 2 sec. following the onset of the CS. There were no indications of inhibition of delay. Possible methodological improvements were discussed.

References

Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.

Brown, J. S., Kalish, H. I., & Farber, I. E. Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *J. exp. Psychol.*, 1951, **41**, 317–328.

Church, R. M., & Black, A. H. Latency of the conditioned heart rate as a function of the CS-UCS interval. *J. comp. physiol. Psychol.*, 1958, **51**, 478–482.

Gerall, A. A., & Woodward, J. K. Conditioning of the human pupillary dilation response as a function of the CS-UCS interval. *J. exp. Psychol.*, 1958, **55**, 501–507.

Kalish, H. I. Strength of fear as a function of the number of acquisition and extinction trials. *J. exp. Psychol.*, 1954, **47**, 1–9.

Kimble, G. A., & Ost, J. W. P. A conditioned inhibitory process in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 151–156.

Libby, A. Two variables in the acquisition of depressant properties by a stimulus. *J. exp. Psychol.*, 1951, **42**, 100–107.

McAllister, W. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–422.

Meryman, J. J. The magnitude of startle response as a function of hunger and fear. Unpublished master's thesis, State University of Iowa, 1952.

MERYMAN, J. J. The magnitude of an unconditioned GSR as a function of fear conditioned at a long CS-UCS interval. Unpublished doctoral dissertation, State University of Iowa, 1953.

MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38,** 89–101.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48,** 162–166.

MURFIN, F. L. The relationship of fear to the CS-UCS interval in acquisition and the CS-duration in extinction. Unpublished doctoral dissertation, State University of Iowa, 1953.

NOTTERMAN, J. M., SCHOENFELD, W. N., & BERSH, P. J. Conditioned heart rate response in human beings during experimental anxiety. *J. comp. physiol. Psychol.*, 1952, **45,** 1–8.

REYNOLDS, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35,** 15–30.

ROSS, L. E. Conditioned fear as a function of the CS-UCS and probe stimulus intervals. Unpublished doctoral dissertation, State University of Iowa, 1959.

SMITH, K. Conditioning as an artifact. *Psychol. Rev.*, 1954, **61,** 217–225.

SPENCE, K. W., & RUNQUIST, W. N. Temporal effects of conditioned fear on the eyelid reflex. *J. exp. Psychol.*, 1958, **55,** 613–616.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.

SPOONER, A., & KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60,** 321–344.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the time of delay. *J. exp. Psychol.*, 1952, **43,** 357–362.

WOLFLE, H. M. Conditioning as a function of the interval between the conditioning and the original stimulus. *J. gen. Psychol.*, 1932, **7,** 60–103.

ZEAMAN, D., DEANE, G., & WEGNER, N. Amplitude and latency characteristics of the conditioned heart response. *J. Psychol.*, 1954, **38,** 235–250.

(Received March 3, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 274–282

THE PERCEPTION OF ALL PATTERNS PRODUCED BY A SEVEN-LINE MATRIX[1]

E. T. KLEMMER

IBM Research Center

The experiments described below were designed to gain insight into the reasons why certain patterns are easier or more difficult to see than others, and to determine the importance of training and experience in the perception of such patterns; e.g., the years of experience with arabic numerals and the several hours of training with other classes of patterns of similar complexity. The total class considered was a set of patterns logically related to the class of numerals. A pattern-source, just complex enough to produce a reasonable set of arabic numerals, was built, and the total class of patterns capable of being produced by this source was studied.

Any display matrix that is capable of producing all the arabic numerals by various combinations of its elements will also be capable of producing many more figures, most of which are not arabic numerals. For example, a 4×5 dot matrix often used for forming numbers is capable of forming 2^{20} or more than a million patterns. The 7-element straight-line matrix used in the present study is capable of only 2^7 or 128 different patterns all of which were evaluated in the study of pattern difficulty.

In a second experiment, performance was measured using a post-stimulus cue asking for only a partial response. This condition reduces the interfering effects of elaborate responses and provides a more direct measure of perceived information. With regard to the question concerning the difference between perceived and transmitted information, if it can be shown that *S* perceives much more information than he transmits with a given response code, the hope is immediately present that with proper recoding this transmission score would approach his perception score.

Apparatus and Common Method

The pattern source was a straight-line matrix device, which has often been used for the production of numerals. In the present study it consisted of a light box containing seven small incandescent bulbs, each of which illuminated one straight-line segment of diffusing material on the front of the box. When all seven bulbs were lit, a figure 8 was formed. Each figure 8 was 4 in. high and 2½ in. wide. Four boxes were placed in a rectangular pattern with 1-in. spaces between patterns, or, for Exp. II in a straight line with 1-in. spaces between patterns.

Figure 1 shows the appearance of the rectangular stimulus field with a typical random pattern of lights. The *S*s were seated 20 ft. from the stimulus field for the rectangular array and 34 ft. for the linear array so that the horizontal visual angle subtended was approximately 2°. The lights were programed by IBM cards and a reader capable of reading 28 holes simultaneously. The current was applied to the circuit for .08 sec. to produce the tachistoscopic exposure. The incandescent time to 90% light output for the bulbs is .04 sec. so that the actual exposure was only about .04 sec. The flash was clearly visible in the dimly lighted room.

The *S*s responded by attempting to duplicate the stimulus pattern on specially pre-

[1] Some of the experiments described in this paper were performed at the Operational Applications Laboratory of the Air Force Cambridge Research Center (Klemmer & Loftus, 1958) and others at the IBM Research Center (Klemmer & Kaplan, 1959). The contribution of Joseph Loftus and Ira Kaplan who conducted parts of the experiments described here is gratefully acknowledged.

pared answer sheets. The answer sheets were printed with a dotted outline of block figure "8s" in the same format as the stimulus. The *Ss* used the dotted outlines to guide their reproductions of the patterns.

The performance of the *Ss* was scored by two methods. First, the number of individual line segments wrong was counted with no distinction between added and omitted lines. Second, each 7-element pattern was scored as correctly reproduced or in error regardless of the number of errors. The line-segment score tends to favor the patterns containing a small number of lines, since *S* often leaves blanks when he is uncertain. The perfect reproduction method gives no credit for close misses. Despite these differences, the major findings were about equal for both scoring methods. In the case of pattern difficulty, both scores are used. Estimates of the amount of information perceived are made in Exp. II but sampling problems prohibit the general use of this measure.

Experiment I

Method

Subjects.—Four college students without previous training served in Exp. I.

Procedure.—Each *S* performed separately and controlled his own presentation rate. When he was ready for the exposure, *S* would look at the light boxes and push the exposure button which flashed the stimulus pattern. He would then reproduce the pattern as best he could by drawing pencil lines on the dotted figure 8 outlines. After operating another switch which turned on the stimulus pattern continuously for scoring, *S* would mark the individual line segments in error (both omission and commission) in all four positions, turn off the scoring lights, advance the card reader to the next stimulus, and be ready to repeat the cycle.

This procedure was repeated until 72 exposures had been presented in one "session." In general, one session was given per day for five days each week with a break of several weeks between the two parts of the experiment described below. Each *S* monitored his own improvement throughout the sessions by noting the number of exposures in which he had no pattern in any of the four positions with more than two line segments wrong. This method of scoring was used to direct attention to all four positions. Earlier tests indicated that the *Ss* had a tendency to watch only two of the four positions and guess at the other two.

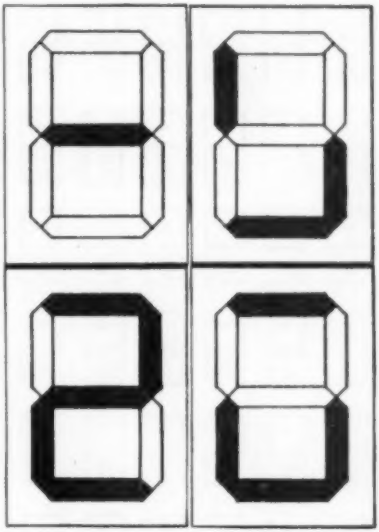

Fig. 1. Appearance of the stimulus field in Exp. I with typical random pattern of lights indicated by black areas. In Exp. II the four patterns appeared in line.

In all sessions, all possible patterns and all combinations of the four patterns were equally likely to appear in each position. No control was exercised over the total number of appearances of each pattern or its location among the four boxes. A different random order of presentation was used in each session.

Results

To get reasonably reliable measures of relative pattern difficulty the performance on each of the 128 patterns was averaged over all 24 sessions and all 4 *Ss*. The score found for each pattern by this method is the average of 216 scores on the average, where each score is the number of correct elements out of a possible seven or the binary correct-incorrect score. The mean difficulty scores are not as stable as might be expected from the average of 216 scores since even the most difficult

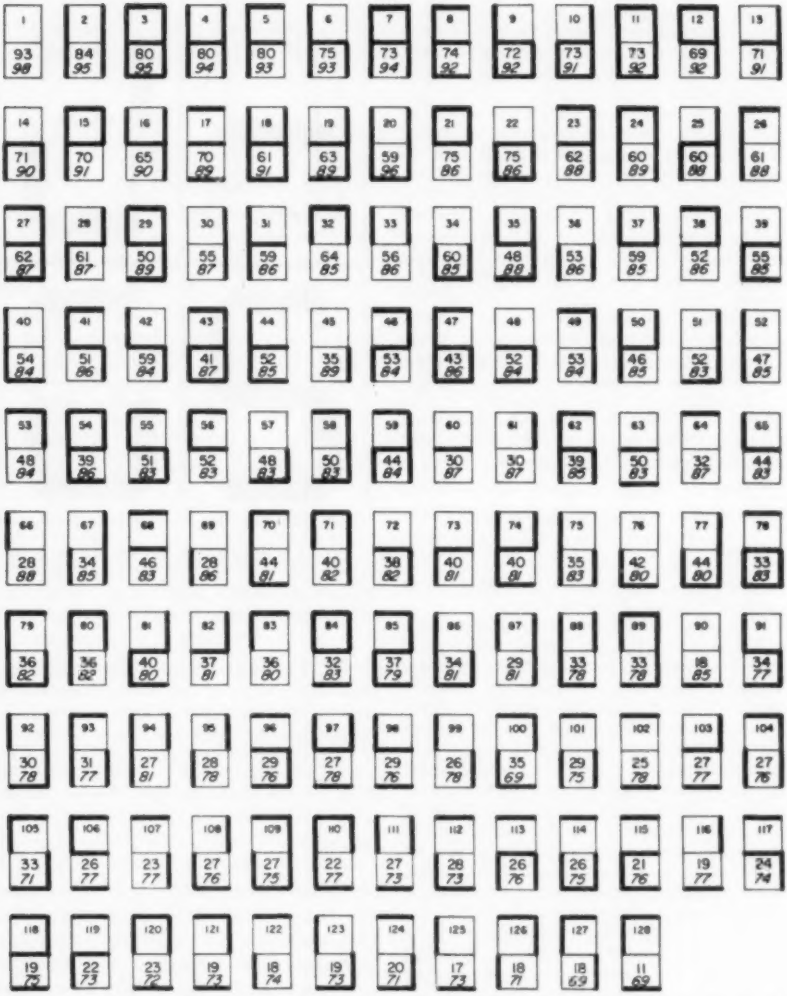

FIG. 2. Order of difficulty of the 128 patterns. Ranks as indicated by the top number in each pattern is based on the average of ranks by pattern-correct percentage (middle number) and line-segment percentage (bottom number). The data are from 27,648 patterns shown to each of four *S*s in Exp. I.

patterns are reported almost perfectly if they occur next to a blank or near-blank.

Figure 2 shows the rank order of difficulty of the 128 patterns based on average scores for all 4 *S*s and all 24 days. The number in the top of each pattern is the rank based on the average rank between the line segment and pattern-correct scores.

The middle number of each pattern is the percentage of presentations which resulted in perfect reproduction; the number in the bottom of each pattern is the percent of line elements correctly reproduced for that pattern.

A note of statistical caution must be used in the interpretation of Fig. 2. Because of the variability in scores, a large difference in rank position between two patterns must exist before the difference can be considered reliable. In order to be 95% sure that the difference in difficulty is real, the patterns must be at least 30 rank positions apart. This figure is based on a comparison of ranks of mirror-image patterns and on an estimate of the error variance of the rank difference score based on rank differences between *Ss*.

In order to determine the stability of the order of difficulty during training for the four *Ss*, the average line-segment score for each pattern based on the first 12 sessions was correlated with scores based on the last 12 sessions. If the data of all four *Ss* are combined, this correlation is .84; however, one *S* (JT) showed a a correlation of only .26 between the first and last 12 sessions (his overall ranking correlated .53 to .60 with the other three *Ss*). Thus, some change in relative difficulty with training is confounded in the combined data but the analysis of Exp. I and II in the Discussion section shows that exactly the same conclusions about what makes patterns easy to see obtain from these different *Ss* and a different training procedure.

EXPERIMENT II

Method

Subjects.—Ten college students, all different from Exp. I, served as *Ss* in Exp. II. All had training in a previous tachistoscopic scale-reading study.

TABLE 1

ORDER, DESCRIPTION, AND RESULTS FROM 10 *Ss* IN EXP. II

Day	Description	Cue Delay (Sec.)	Line Segments Correct %
1	Practice: partial R, auditory cue	uncontrolled	
2	Practice: partial R, visual cue	0.5	73
3	Practice: partial R, visual cue	0.1	75
4	Partial R, visual cue	0.1	80
5	Partial R, visual cue	0.4	79
6	Partial R, visual cue	1.6	78
7	Partial R, visual cue	3.2	78
8	Partial R, visual cue	0.8	84
9	Partial R, visual cue	0.2	89
10	Complete R	no cue	72
11	Complete R	no cue	74
12	Complete R	no cue	77
13	Partial R, visual cue	0.2	90
14	Partial R, visual cue	0.8	87
15	Partial R, visual cue	3.2	84
16	Partial R, visual cue	1.6	87
17	Partial R, visual cue	0.4	92
18	Partial R, visual cue	0.1	91
19	Partial R, auditory cue	0.1	95
20	Partial R, auditory cue	3.2	87
21	Partial R, visual cue	0.1	95
	Partial R, visual cue	3.2	89

Procedure.—All 10 *Ss* were tested simultaneously as a group viewing the linear arrangement of stimulus patterns. A warning buzzer sounded 1 sec. before the tachistoscopic flash. After all *Ss* had recorded their responses, the stimulus pattern was turned on and left on while the *Ss* scored their papers. Test sessions lasted about 1 hr. on each day and contained 64 presentations for the poststimulus cued conditions and 42 or 43 presentations for the complete-response condition.

For the cued conditions, the *Ss* were instructed to reproduce only that one of the four patterns denoted by the poststimulus cue. The poststimulus cue, instructing which position to report, consisted of a dim red light that appeared below the critical position. The time between stimulus and cue was accurately controlled by an RC delay circuit. Three practice sessions were given as indicated in Table 1, followed by sessions in which the cue delay was varied systematically in log steps from .1 sec. to 3.2 sec. in order to measure the decrement due to time only. The order of delay times is given in Table 1. The longest delay (3.2 sec.) is roughly half the time taken in record-

ing all four positions so that the difference in performance at the longest delay in the sampled perception tests and the performance in reporting all four positions is a measure of the decrement caused by recording the additional positions. Three sessions in the middle of the design required *Ss* to report all four patterns in fashion similar to Exp. I.

In addition to the visual poststimulus cue, two tests were conducted with an auditory cue. The auditory cue consisted of a spoken digit, *one*, *two*, *three*, or *four*, calling for *Ss* to report the corresponding position. The auditory cues were recorded on magnetic tape. The onset of the stimulus was also controlled by the tape reader so that accurate control over the delay time was exercised. Only the extreme delays (.1 and 3.2) were tested with the auditory cue. Control visual cue sessions at .1 and 3.2 sec. were run immediately following the auditory sessions.

Each of the 128 patterns appeared once in each of the four positions for each condition. The order of patterns and the grouping of patterns was random and changed between sessions. Each of the four positions was called for by the cue an equal number of times for each delay time, but the order of positions was irregular. Each delay was run on two days from Days 4 through 18, separated in a symmetrical design as shown in Table 1, with half of the 128 patterns shown on each day in the critical position.

In scoring the responses, each of the seven possible lines of the pattern in the critical position was scored separately, right or wrong, so that the total scores for each response varied from 0 to 7. Responses were scored immediately by *Ss* and checked later for purposes of data analysis.

In order to get some idea of the verbal associations involved with the entire set of 128 patterns, the following controlled association test was run with the 10 *Ss* of Exp. II. Each of the 128 patterns was shown once in random order in the usual manner using only a single light box. The *Ss* were instructed first to write down as many names for the pattern as came to mind, then reproduce the pattern on the answer sheet, and finally, rate it for ease or difficulty on a 5-point scale with 5 assigned to the easiest patterns and 1 to the most difficult.

Results

Average performance for the 10 *Ss* for each session is given in Table 1 in terms of the percentage of line segments correct. The main body of sessions in this experiment were run in a symmetrical design with the hope that the learning effect would be small enough to allow an assumption of linear improvement with test days and a simple averaging of performance on repeated runs of the same test. This method of eliminating the learning effect would not work because of the large nonlinear change with practice, but fortunately another method of averaging does take account of the nonlinear learning curve. This method is simply to take the geometric mean of the error scores of repeated runs on the same test. If the assumption underlying the method is met, this geometric mean will give the score that would have been obtained if the test had actually been run on a day midway between the actual days. For the symmetrical design covering Test Days 4 through 18, all the test will be corrected to Day 11. The assumption underlying this method of averaging is that improvement with practice follows a negative growth function asymptotically approaching perfect performance. In view of the high percentage test scores on the last days, this assumption seems very reasonable.

The line-segment scores from the six delays of the partial response tests are corrected to Test Day 11, as explained above, and are plotted in Fig. 3. Note that the delay times along the abscissa are plotted on a logarithmic scale.

The best performance occurs in the partial response test with .2-sec. delay between stimulus exposure and the cuing light. Performance drops off with delays on either side of .2 sec., but the drop off is not rapid. It is quite possible that the drop off in performance at the shortest delay was

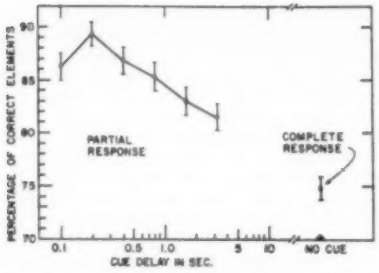

FIG. 3. Average performance for 10 *S*s in Exp. II as a function of time between stimulus and poststimulus cue. Vertical lines denote 90% confidence intervals.

due to interference between stimulus and cue which were both visual. Unfortunately, since these results were not known when the auditory cues were treated, the .2-sec. delay was not used with the auditory cue.

The confidence intervals for the partial response points in Fig. 3 are based on the variance due to patterns of the combined score of all 10 *S*s and thus refer to these 10 *S*s only. In the case of the complete response point each pattern occurred four different times, once at each position. The confidence interval is based on the variance over these four presentations of the combined scores of all *S*s.

Partial response tests with auditory cues were run on Days 19 and 20, followed on Day 21 with repeat visual cue tests using the same delays. Table 1 shows that the short delay (.1 sec.) developed a 95% line-segment score in both modalities whereas the long delay (3.2 sec.) showed 87% correct with the auditory cue and 89% correct with the visual cue. The difference between 89% and 87% may well have been due to practice. Indeed, an actual difference between the auditory and visual cue scores at the short delay may have been hidden by an equal and opposite practice effect. At any rate, the conclusion must be that any difference between auditory and visual cuing is small.

Information Measures

In addition to the percentage-correct measures, *S*'s performance can be measured in terms of the amount of information transmitted or perceived. We shall distinguish between transmitted and perceived information in the following way. That part of the input information that is also in *S*'s response is transmitted information. If *S* is asked to make discriminative responses to the entire stimulus, the transmitted information score will be the only one used. If *S*'s perception of the stimulus is only sampled—by providing a random poststimulus cue to respond to only part of the stimulus—then it will be assumed that *S* perceives and temporarily stores an amount of information equal to the transmitted information multiplied by the sampling ratio. In the partial response tests of the present study, the sampling ratio is always four because the total stimulus contains four times as much information as the part asked for by the poststimulus cue.

A straightforward application of the usual formulas for computing information transmission is not possible because of the large number of alternatives involved. There are 268,435,456 possible different stimuli when all four light boxes (positions) are used with all possible patterns. To reduce the number of alternatives to a workable number, each position was analyzed separately, which means that interaction between positions was neglected. In addition, each *S*'s responses to all patterns of the test were combined into a single frequency distribution according to number of errors. Also, in

TABLE 2

ESTIMATED INFORMATION PERCEIVED IN EXP. II

Visual Cue Delay (Sec.)	Mean Test Day	Aver. *S* (bits)	Best *S* (bits)
0.1	11	14.2	22.5
0.2	11	17.2	23.3
3.2	11	11.9	20.0
Complete R	11	9.6[a]	13.2[a]
0.1	21	21.6	26.4
3.2	21	17.3	24.0

Note.—Perfect performance is 28 bits in all tests.
[a] Information actually transmitted.

estimating probabilities of incorrect response patterns, all response patterns with an equal number of errors were assumed equiprobable. For the complete response tests, separate calculations were made for each position, but for the partial response tests, positions were not analyzed separately. This method of calculating informational measures leads to an underestimate of the information transmitted because it averages out factors that would reduce response uncertainty.

The lower bound for information transmission and that for information perceived, calculated as described, are given in Table 2. The scores for the best *S* were from the same student in all tests except the numerals where he was excelled by another *S*.

The highest informational values occurred in the partial response tests in which *S* was not burdened with the necessity of transmitting all of the information to his answer sheet. The rerun of the partial response test with .1-sec. cue delay shows that the average *S* perceived almost 22 bits. This is equivalent, informationally, to seeing seven numerals almost perfectly.

Augenstine and Wayner (1956) have tried a partial response with a short flash of capital letters and found best perception at four to six letters. The average information perceived for their three *S*s was 16.0 bits for four letters and 16.2 bits for six letters. Numerals would almost certainly show poorer informational scores under the same conditions so it seems unlikely that the digit code could ever get to the 22 bits found with the class of 128 patterns in the partial response test with .1-sec. cue delay.

Table 3 gives the percentage of the total information transmitted by position for the complete response tests. The scores for all 10 *S*s show that the left positions contribute more than the right positions, and that in the complete response 4-position test the middle positions contribute more than the ends. Quite another finding emerges on this same complete response 4-position test when the data from individual *S*s is studied. This breakdown is also shown in Table 3. Four *S*s did best on the left position, three *S*s on the middle position, and two *S*s on the right position. One *S* did well on the two ends and very poorly on the middle positions. In this test, in which it was impossible for him to see and report all positions, *S* apparently chose to concentrate his attention on two or three positions and let the other position(s) go.

TABLE 3

PERCENTAGE OF TOTAL INFORMATION TRANSMISSION CONTRIBUTED BY EACH POSITION IN THE COMPLETE RESPONSE CONDITION OF EXP. II

Subjects	Position			
	1	2	3	4
All 10 *S*s	22	38	26	15
Four left-oriented *S*s	44	35	17	4
Two right-oriented *S*s	5	32	30	33
Three middle-oriented *S*s	2	56	39	3
One ends-oriented *S*	51	4	3	42

Most of the *S*s in the 10-*S* group gave one or more names in response to nearly all of the 128 single-position patterns presented in the association test. Typical names were "Mutt and Jeff, backward C, magnet, question mark, staircase, chair." The number of associative responses given by all *S*s to each pattern correlated .47 with pattern ease as measured by average line-segments-correct score. A correlation with performance was also obtained by counting the number of *S*s giving the same response to each pattern. This frequency of the most common response correlated .37 with the average line-segments-correct score. Finally, the mean ease-of-seeing rating for each pattern by all *S*s correlated .57 with the line-segment-correct score.

DISCUSSION

What Makes Patterns Easy to See?

Let us now look at the order of difficulty in Fig. 2 and see what characteristics relate to pattern difficulty. The most important aspect of patterns which relate to difficulty is the continuity of the figure. Continuous patterns are seen perfectly 52% of the time, while discontinuous patterns are seen perfectly only 30% of the time (Exp. I). Because of the importance of this one aspect, only continuous patterns are included in comparing other characteristics in order to avoid confounding the comparison with relation to pattern continuity. Table 4 gives the average patterns-correct and line-segment scores for several of the more obvious categories of patterns for Exp. I. Only categories leading to differences significant at the .05 level or better are included. The line-segment scores obtained by averag-

TABLE 4

MEAN PATTERN DIFFICULTY SCORES IN TERMS OF PERCENTAGE CORRECT FOR SEVERAL CLASSES OF PATTERNS

Description of Patterns	No. of Patterns in Class	Exp. I		Exp. II
		Patterns-Correct Score	Line-Segment Score	Line-Segment Score
All patterns	128	44%	82%	81%
Discontinuous patterns	47	30	77	78
Continuous patterns	81	52	86	83
Symmetrical, both axes	5	71	93	85
Symmetrical, vertical axis	12	59	88	83
Asymmetrical	56	48	85	80
Containing closed loop	14	64	90	85
Without closed loop	67	49	85	80
Letters and numerals	39	56	87	84
Not letters or numerals	42	48	85	82

ing over all conditions of Exp. II are also presented in Table 4 to indicate the reliability of the category scores. Since the linear arrangement of patterns in Exp. II was viewed at twice the distance as the rectangular arrangement of Exp. I each pattern subtended only half the visual angle. This difference and the use of a partial response in most of Exp. II might be expected to produce some differences in relative difficulty but it is heartening that there are no differences in order of difficulty of the pattern classes in Exp. I and Exp. II as shown in Table 4.

It is interesting to note that performance on letter and numeral forms was only slightly better than on the continuous forms, which are not letters or numerals. The more abstract classifications of continuity, symmetry, and presence of a closed loop are much more important determiners of ease of seeing. Patterns symmetrical only about the horizontal axis were not seen as well as patterns symmetrical about the vertical axis. They were seen better than asymmetrical patterns, but the differences were not statistically significant. Fitts, Weinstein, Rappaport, Anderson, and Leonard (1956) report that patterns symmetrical about the horizontal axis only are not seen as well as patterns symmetrical about the vertical axis.

There are many other ways of classifying the patterns, but none that we have tried seem intimately concerned with pattern difficulty. It might appear that the number of lighted line-segments is related to pattern difficulty but actually there is no consistent relation between score and number of lines except that which might be predicted from continuity, symmetry, and letter or numeral shapes. It is important to bear in mind that four patterns were presented simultaneously and, therefore, the interaction between adjacent patterns has a strong influence on performance. There is no obvious difference, however, between the horizontal linear arrangement of patterns used in Exp. II and the rectangular arrangement in Exp. I.

How Important Is Familiarity?

The complexity of the results of the present study may tend to hide the general significance of the findings. Perhaps the most general finding is that the familiar numerals and letter forms do not stand out like the proverbial

sore thumb when shown with nonsense figures of similar geometric properties. Rather, some abstract classifications are even more important correlates of pattern difficulty.

In addition to being familiar forms, the numerals and letters have well-learned names and so the research on verbal labels and discrimination is pertinent. Both positive (Hake & Erickson, 1956) and negative (Robinson, 1955) evidence has been found for the contribution of verbal labels to discrimination. A possible explanation for this disagreement is that verbal labels facilitate discrimination when only part of the stimulus information is pertinent (i.e., the stimuli are redundant). If all of the stimuli that the source is capable of generating are used and all differences between them are important, the use of verbal labels may result in no better performance than straight reproduction (Arnoult, 1956).

Summary

Two experiments were conducted which required the reproduction of linear visual patterns following brief exposures. The patterns were produced by randomly selecting lines from a 7-line matrix in the form of a block figure 8. This source is capable of generating a reasonably good set of arabic numerals and 118 other patterns.

Four matrices were presented simultaneously and programed to produce randomly selected patterns from the entire class. Some tests utilized a poststimulus cue which denoted only one of the four patterns for reproduction while other tests required reproduction of all four patterns.

Performance was measured both in terms of line segments reproduced correctly, patterns reproduced correctly and amount of information perceived. The following conclusions are for the combined data of the number of *Ss* indicated and for moderate amounts of training. All percentages refer to the line-segment scores:

1. A delay of .2 sec. between stimulus and poststimulus visual cue gave best performance. The average score was down 7% with a delay of 3.2 sec. A check of two delay times using an auditory cue showed no difference between visual and auditory cuing (10 *Ss*).

2. Performance when a complete report of the four matrices was required was 14% below performance with the partial response and a .2-sec. cue delay (10 *Ss*).

3. The arabic numerals were reported only slightly better than nonsense figures when all patterns were mixed. Continuous patterns were seen better than broken patterns. Patterns containing a closed loop of line segments were seen better than patterns without such a loop. Symmetrical patterns were seen better than asymmetrical patterns. Patterns judged meaningful were seen better than those not meaningful and the blank pattern was seen best of all (14 *Ss*).

4. In a controlled association test, the total number of associations given to each pattern by all *Ss* correlated .47 with a performance measure (10 *Ss*). An average rating score of pattern ease by *Ss* correlated .57 with a performance measure (10 *Ss*).

5. An informational analysis of Exp. II showed that the average *S* transmitted 9.6 bits when the complete response was required, but perceived 17.2 bits as estimated from a partial response test with optimum cue delay (21.6 bits after additional training) (10 *Ss*).

References

Arnoult, M. A comparison of training methods in recognition of spatial patterns. *USAF Personnel Train. Res. Cent. tech. Note*, 1956, No. AFPTRC-TN-56-27.

Augenstine, L., & Wayner, M. Human performance in information transmission: IV. Flash recognition of familiar displays. *Control Sys. Lab. Rep.*, 1956, No. R-69.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, **51**, 1956, 1–11.

Hake, H. W., & Ericksen, C. W. Recognition and identification of complex visual forms as a function of the labeling system employed. *USAF WADC tech. Rep.*, 1956, No. 55–367.

Klemmer, E. T., & Kaplan, I. The perception of patterns from a seven-line matrix. *IBM Res. Cent. Rep.*, 1959, No. RC-124.

Klemmer, E. T., & Loftus, J. P. Numerals, nonsense forms and information. *USAF Cambridge Res. Cent. tech. Rep.*, 1958, No. TR 57-2.

Robinson, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, **49**, 1955, 112–114.

(Early publication received July 11, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 283-289

PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF PERCENTAGE OF OCCURRENCE OF RESPONSE MEMBERS AND OTHER FACTORS

HARDY C. WILCOXON, WARNER R. WILSON,[1] AND DALE A. WISE[2]

University of Arkansas

Previously reported studies have shown that rate of learning of paired associates increases directly with percentage of occurrence of response members, hereafter called percentage ORM. Goss, Morgan, and Golin (1959), using paired nonsense syllables at rates of 2:2 sec. and 3:3 sec., found that number of trials and number of correct responses to criterion varied inversely with percentage ORM in line with expectations from the partial reinforcement literature, but attempts to simulate extinction were largely unsuccessful. Schulz and Runquist (1960) found that acquisition of paired adjectives at the 2:2-sec. rate was moderately retarded as percentage ORM varied over 100%, 80%, 60%, 40%, and 20%, but that retention was unaffected when measured by recall and relearning after 23.5 hr.

The three experiments of the present report, all run before the appearance of the studies cited, deal with additional variables, as well as those already reported, and thus provide some new information, some confirmation of old, and a few points of disagreement.

EXPERIMENT I

The purpose of the first experiment was to study the effect on both learning and retention of percentage ORM together with two other variables which might produce interactions, namely, level of ability of *Ss* and degree of original learning.

Method

The general method in all three experiments involved only one notable departure from the standard memory-drum presentation of paired associates by the anticipation method. Lists were prepared so that instead of the response member invariably following the stimulus member on every trial, it was occasionally omitted in some groups, more or less at random for each pair, and replaced by a dashed line.

Subjects.—The *Ss* were 72 general psychology students selected from the top one-half of the various classes they came from. This selection was expected to help insure that all *Ss* would learn within the arbitrary time limit of 1 hr. imposed by class schedules.

Design.—A $3 \times 3 \times 2$ factorial design was employed. Percentage ORM varied in three ways, 100%, 75%, and 50%; level of ability of *Ss* (LA) was varied by subdividing *Ss* into three groups whose general psychology objective test scores placed them in the upper, middle, or lower one-third of the upper one-half of their classes (cf. above); and acquisition criterion (AC) was varied by running half of the *Ss* to a criterion of one errorless trial and half to a criterion of three successive errorless trials. The *Ss* were assigned to subgroups at random before they appeared for the experimental session, except for the restriction that equal numbers of high-, middle-, and low-ranking students were assigned to each of the other subgroups.

Apparatus and lists.—A list of seven nonsense syllable-adjective pairs was learned by all *Ss* on a Hull-type memory drum. The list was taken from a study by Feldman and Underwood (1957) where it had been shown to be a fairly easy list to learn. The nonsense syllable stimulus members had low intralist similarity, and association values ranged from 0% to 20% according to Glaze's (1928) calibration, with a mean association value of 9.5%. Response members were common adjectives with low similarity. The

[1] Now at the University of Hawaii.

[2] Now attached to CNAVNTRA Staff, Naval Air Station, Corpus Christi, Texas.

pairs were typed on white paper tapes in eight different orders, presented in rotation to eliminate serial effects. Three sets of the eight orders were prepared to provide for the different ORM conditions. The tapes presented to 100% ORM groups were standard, the stimulus member always being followed by the stimulus and response members paired. The tapes for 75% and 50% ORM groups, however, had the response member of a given pair replaced by a dashed line on 25% and 50% of the trials, respectively. This omission of the response word was accomplished at random with the following restrictions. In the 75% ORM condition, no more than one stimulus syllable with the response word omitted was allowed in succession within a trial, and a given stimulus member was not allowed to appear without its response word more than one trial in succession. In the 50% ORM condition, no more than two stimulus syllables with the response word omitted were allowed in succession within a trial, and a given stimulus syllable was not allowed to appear without its response word more than two trials in succession. Percentage ORM for a given pair averaged to the exact value for each condition over blocks of eight trials and was kept as close as possible to the stated values within individual trials.

In addition to the main lists, two-pair warm-up lists of the same general type were prepared for each percentage ORM.

Procedure.—The *Ss* received instructions read verbatim by *E* which explained the anticipation method and, where appropriate, took note of the fact that the response word would sometimes be replaced by a dashed line on the tape. Each *S* then learned the warm-up list having the same percentage ORM as he would encounter in the main task.

The main task was then presented at a 2:2-sec. rate, with approximately 15 sec. elapsing between trials. After *Ss* met the appropriate criterion *E* dismissed them with no indication that they would later be asked to try to recall what they had learned. Six to eight days later, however, each *S* was called back for recall and relearning tests. Recall scores were obtained by giving *Ss* two trials at the drum with 0% ORM lists. Immediately following the recall test, all *Ss* relearned the seven pairs to a criterion of one errorless trial, using 100% ORM lists.

Results and Discussion

The results of this experiment can be summarized succinctly by saying that they were all negative. They are important only because there seemed excellent reason to believe some of them would be positive.

Separate analyses of variance dealing with trials to a criterion of one perfect repetition, average number of words recalled in the two recall trials, and number of relearning trials to a criterion of one perfect repetition all yielded negative results, with F ratios of 1.05, .75, and .47, respectively. Data on the three levels of ORM, although not statistically significant, are of some interest, and are presented in Table 1 where differences in levels of ability and degree of learning are ignored inasmuch as the latter two variables were apparently not effectively induced.

Hindsight provides a variety of rationalizations for the failure of LA and LA × ORM to achieve significance, but the nonsignificant effect of ORM itself was quite surprising. This finding receives some confirmation, however, in the Schulz and Runquist (1960) data, where differences in mean trials to one perfect repetition among 100%, 80%, and 60% ORM groups were small and nonsignificant by separate t tests calculated by the present authors. The significant F over all five of their groups seems to have arisen primarily from the reduction of percentage ORM to 40% and 20% in the remaining two groups.

The Goss, Morgan, and Golin (1959) data, however, present a more complicated picture. They found large and significant differences in acquisition among the 100%, 75%, and 50% ORM groups in their experiment run at a 2:2-sec. rate of exposure, the rate used in the present study. However, their materials were eight nonsense syllable pairs, while ours were seven nonsense syllable-adjective pairs; and their analysis of acquisition data was based upon a criterion of seven out of eight pairs correct rather than one perfect repetition. These differences in method may account

TABLE 1

MEANS AND *SD*s IN ACQUISITION, RECALL, AND RELEARNING IN EXP. I

% ORM	Acquisition Trials (1 Perfect)		Recall Trials 1 + 2/2		Relearning Trials (1 Perfect)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
100%	20.9	9.4	3.6	1.7	3.8	2.3
75%	20.8	11.3	4.1	1.8	3.7	2.3
50%	22.5	9.7	4.1	1.7	4.0	3.0

for the discrepancy in results. In their experiment run at a 3:3-sec. rate, however, differences among the 100%, 75%, and 50% groups were smaller and not significant by separate t tests based on σ_m calculated from the within groups variance estimate, which is in line with the present results.

EXPERIMENT II[3]

The surprising failure to find a significant retardation of acquisition in Exp. I with ORM as low as 50% led to Exp. II, the main purpose of which was to reduce ORM to a lower percentage in search of the retardation in acquisition which the logic of the situation would seem to require. A second purpose was to investigate the effect of reversing the S-R order of the NS-Adjective pairs in an effort to shed light on the reasons behind the ineffectiveness of reduced percentage ORM in Exp. I. The S-R order, Adjective-NS is known to be harder to learn than the order, NS-Adjective (Kimble & Dufort, 1955; Noble & Stockwell, 1957). This suggested the possibility of an interaction between S-R order and ORM.

Method

Design.—A 3 × 2 × 3 factorial design was utilized. Percentage ORM varied in three ways, 100%, 50%, and 25%; S-R order of the pairs (O) varied in two ways, NS-Adjective, and Adjective-NS; and three comparable sets of paired-associate lists (L) were employed merely to provide a more adequate sample of learning material.

Subjects.—Ninety undergraduate psychology students served as *S*s and were randomly assigned to the 18 conditions generated by the design.

Apparatus and lists.—The apparatus was the same as that used in Exp. I. The three seven-pair lists consisted of nonsense syllables taken from Glaze's (1928) list of syllables having 0% association value, and adjectives from Haagen's (1949) lists showing high familiarity. The various memory drum tapes required by the design were prepared exactly as in Exp. I. In this case they represented eight orders of each of the three lists of seven pairs, two S-R orders of the pairs, and three ORM conditions. The restrictions on purely random omission in the 25% condition were that no stimulus member was allowed to appear on more than four successive trials without its response member, and no more than four successive response-member omissions were allowed within a trial.

Procedure.—The procedure for the acquisition phase of Exp. I was followed exactly. The criterion was one errorless trial. Both nonsense syllables and adjectives were pronounced aloud when anticipated.

Results and Discussion

Subgroup means and *SD*s, with lists ignored, are presented in Table 2. Table 3 shows that analysis of variance yielded significant effects for percentage ORM and S-R order (O) but, surprisingly, interaction of ORM × O did not occur. All t tests between means of groups differing

TABLE 2

MEANS AND *SD*s OF TRIALS TO CRITERION IN EXP. II

S-R Order	100% ORM		50% ORM		25% ORM	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
NS-Adjective	17.6*	7.9	20.4	8.8	31.8	12.8
Adjective-NS	26.3	7.7	31.4	7.8	42.5	14.6

* σ_m based on within-groups variance estimate was 2.77. N = 15 for subgroup means.

[3] We wish to acknowledge the assistance of James J. McRoy in running the *S*s and performing the statistical analysis of this experiment.

TABLE 3

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION OF ONE ERRORLESS TRIAL IN EXP. II

Source	*df*	*MS*	*F*
%ORM	2	1,866.43	16.32*
S-R order (O)	1	2,340.90	20.47*
Lists (L)	2	1.63	
Total interaction	12	108.32	
Within groups	72	115.36	
Total error	84	114.36	

* $P < .01$.

only in S-R order are significant ($P < .05$). This is not true, however, of all the mean differences between groups which differ only in percentage ORM. The comparisons between 50% and 25% are highly significant ($P < .01$), but the comparisons between 100% and 50% failed to reach significance ($P > .05$).

An additional analysis of the data was suggested by the Schulz and Runquist (1960) study. They analyzed errors following first correct anticipation of response members to investigate possible differences among the percentage ORM groups. A measure called the Correct/Opportunity ratio was computed for each *S* based on the total number of correct anticipations following the first correct, divided by the number of trials (opportunities) left to criterion. The analysis showed that the Correct/Opportunity ratio was an inverse function of percentage ORM. An identical analysis of the present data, however, failed to confirm this finding. Subgroup means and *SD*s, lists ignored, are shown in Table 4. Analysis of variance over all subgroups, including lists, resulted in an *F* ratio of only .624. The writers are at a loss to explain this discrepancy in results.

The significant *F* ratio for percentage ORM confirms the findings of Goss, Morgan, and Golin (1959) and of Schulz and Runquist (1960) while the significant effect of S-R order of the pairs is in line with the results of Kimble and Dufort (1955) and Noble and Stockwell (1957).

Lack of interaction between percentage ORM and S-R order was surprising and appears to mean that familiarity, as distinct from meaningfulness, is not a powerful variable in this situation. That is, reduced percentage occurrence of nonsense response members (which automatically reduces *S*'s opportunity to become familiar with them) retards learning no more than does reduced percentage occurrence of meaningful adjectives with which *S* is already familiar.

In our view, the most important aspect of these findings is the surprisingly small retardation in learning resulting from moderately reduced ORM. It is obvious that learning must suffer if reduction in ORM is carried far enough. Yet, *S*s in reduced percentage ORM groups apparently get much more benefit from each glimpse of the response member allowed them than *S*s in the conventional, 100% ORM groups. For example, *S*s in 100% ORM groups saw the response member an average of about 22 times before mastery. The *S*s in 50% ORM groups, however, saw it only 13 times, and those in 25% groups only about 9 times before mastery.

Thus, it appears that *S*s in reduced ORM groups utilize the information presented via the memory drum somewhat more efficiently than *S*s under the conventional, 100% ORM condition. The experiment which follows was designed to test some of the implications of this general idea.

TABLE 4

MEANS AND *SD*S OF CORRECT/OPPORTUNITY RATIOS

S-R Order	100% ORM		50% ORM		25% ORM	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
NS-Adjective	.78	.099	.77	.132	.74	.131
Adjective-NS	.78	.086	.76	.084	.76	.093

Experiment III

The planning of this experiment proceeded from a consideration of some of the general ideas of information theory (Shannon & Weaver, 1949; Weaver, 1949).

Specifically, the problem was to explore the possible interactions of ORM, rate of presentation (R), and length of list (L). A possible reason for the relative ineffectiveness of moderately reduced percentage ORM could be that the reduction does not bring the amount of information below the maximum which *S* is capable of assimilating in the unit time involved per trial. A further implication appears to be that "input overload" might occur with high rates of presentation and/or long lists, forcing a general error and confusion (cf. Weaver, 1949, p. 14). Thus, reduced percentage ORM could conceivably be beneficial under some conditions of exposure rate and length of list.

Method

Design.—The design was 3 × 3 × 3 factorial. The three values of each variable were: ORM, 100%, 50%, and 25%; rate of presentation (R), 1:1-sec., 2:2-sec., and 4:4-sec.; and length of list (L), 4, 6, and 8 pairs. It should be noted that the design confounded rate of input and the time allowed for anticipation of the response member.

Subjects.—The *Ss* were 108 general psychology students naive with respect to the experimental task. They were randomly assigned to the 27 experimental conditions. Data from 10 additional *Ss* had to be discarded because of apparatus failure, *E*'s errors, *Ss*' failure to understand instructions, and a reported inability to concentrate on the task. The discarded data came from a variety of the 27 experimental conditions, making it unlikely that results would be biased in consequence.

Materials.—Lists were constructed from 16 nonsense syllables of less that 15% association value drawn from Trapp and Kausler's (1959) Table 1. An attempt was made to select homogeneous syllables which would be relatively easy to articulate as judged by *E*. Other aspects of list construction followed the general rules described earlier except that four rather than eight independent orders of syllable pairs were used, the four orders for each ORM, R, and L condition being typed on single, continuous paper bands. Three-pair warm-up lists were prepared with syllables of the same type used in the main task.

Apparatus and procedure.—One memory drum operating at a 2:2-sec. rate was used for the short practice list. A second drum used for the main task was modified with adjustable idler rollers to accommodate the three different lengths of continuous paper bands required. Following instructions and the warm-up task, *Ss* learned to a criterion of one errorless repetition under continuous trials with no intertrial interval. The nonsense syllables were pronounced when anticipated.

Results

The means and *SD*s for trials to criterion are presented in Table 5 in terms of raw scores. Raw scores were not submitted to analysis of variance, however, since Bartlett's test (Edwards, 1950) revealed that the hypothesis of heterogeneity of variance was tenable. This was remedied by a square-root transformation, and Table 6 gives the corresponding analysis of variance.

F tests revealed significant effects upon trials to criterion for each of the three main variables. Inspection

TABLE 5

Raw Score Means and *SD*s of Trials to Criterion in Exp. III

Length of List	Rate (Sec.)	100% ORM		50% ORM		25% ORM	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
4	1	65.0[a]	31.5	94.2	44.6	108.8	85.7
	2	21.8	4.3	41.0	14.6	49.2	20.6
	4	13.8	5.9	15.8	7.3	35.0	22.6
6	1	77.2	23.4	85.2	53.6	110.0	48.2
	2	37.0	9.3	71.0	22.7	57.0	21.3
	4	18.5	6.4	26.8	5.8	22.2	6.0
8	1	167.0	64.9	215.0	25.8	139.8	34.7
	2	45.5	15.3	64.8	24.7	77.2	20.5
	4	33.8	17.3	24.5	6.8	41.5	17.5

[a] $N = 4$ for individual group means.

TABLE 6

ANALYSIS OF VARIANCE OF TRANSFORMED SCORES OF EXP. III

Source	*df*	Trials		Time	
		MS	*F*	*MS*	*F*
%ORM	2	16.52	5.63*	6.57	6.50*
Rate (R)	2	284.42	96.90*	1.28	1.27
Length (L)	2	53.03	18.07*	83.34	82.36*
%ORM × R	4	2.77		1.50	1.48
%ORM × L	4	1.98		.56	
R × L	4	10.86	3.70*	2.30	2.28
%ORM × R × L	8	3.19	1.09	1.48	
Within	81	2.94		1.01	

* $P < .01$.

of the means reveals that number of trials to criterion varied inversely with ORM and directly with Rate and Length. Most of the expected interactions, however, failed to occur. Percentage ORM did not interact significantly with either Rate or Length but Rate × Length showed a significant interaction, attributable primarily to greatly increased difficulty of eight-pair lists learned at the 1-sec. rate.

Analysis of variance of the time scores, also reported in Table 6, showed that Rate, which exerted a powerful effect in retarding acquisition when trials to criterion was the measure, failed to exert a significant effect when total elapsed time to criterion was the measure of learning. Also, the R × L interaction, which was significant in the analysis of trials to criterion, failed to reach significance with time as the measure.

It was hoped that inclusion of the variables, rate of presentation (R), and length of list (L) in the design of Exp. III would, through their interaction with ORM, assist in understanding the reasons behind the surprisingly small detrimental effects produced by reduced percentage ORM. Since neither interacted significantly with ORM, however, the reasons remain speculative.

Both R and ORM appear to have been varied over a sufficiently wide range to have produced interaction if it exists, but the small number of cases per cell may have precluded detection of a moderate effect. Both variables showed significant separate effects and, in the case of R, the practicable limits of variation were probably approached. The *Ss* in the 1:1-sec. groups appeared to have some difficulty keeping up with the pace of the drum, and the 4:4-sec. rate appeared to be so leisurely as to risk boredom.

Even more caution is warranted in generalizing about the effects of the L variable. The relatively restricted variation of 4, 6, and 8 pairs was chosen primarily because exploratory work showed that longer lists learned at the lowest percentage ORM condition would often require more time than a single free class period allowed, and it by no means exploited the range on the long side.

Thus, the null hypothesis for ORM × R may be accepted only tentatively, and subsequent work with longer lists may well show an ORM × L interaction. Another interaction which seems likely, although not tested for in these experiments, is that of ORM, rate, and meaningfulness or availability at the beginning of learning of the response unit.

The separate effects of R and L upon mean trials to criterion are in general agreement with previously published studies (Hovland, 1938, 1940, 1949). The finding that R has no effect upon total elapsed time to mastery, however, apparently has not been reported previously and is of considerable interest. It shows that *Ss* assimilated the material to be learned at a constant rate which was independent of the rate at which the material was presented to them. Thus, the 1:1-sec. rate apparently did not create an "input overload," although it did appear to force *Ss* near to the limit of their perceptual and articulatory capacities, particularly the latter. Similarly, the 4:4-sec. rate apparently did not reduce the rate of input below *Ss'* rate of assimilation of the material.

Previous investigators (Goss et al., 1959; Schulz & Runquist, 1960) have offered a number of suggestions as to factors which probably operate to compensate for reduced percentage ORM.

Schulz and Runquist were able to show through an analysis of overt errors that response competition was less in paired adjectives learned under reduced ORM. Unfortunately, records of overt errors were not kept in the present studies. Both Goss et al., and Schulz and Runquist note that decreased percentage ORM provides opportunity for increased rehearsal, which is undoubtedly an important factor. Results of the present studies are congruent with both interpretations and we offer the additional speculative suggestion that reduced ORM appears to increase motivation and attention of *Ss*.

Summary

Three experiments on percentage occurrence of response members (ORM) in paired-associate learning were reported. In Exp. I, 72 *Ss* selected from the upper one-half of general psychology classes were used in a factorial design with 100%, 75%, and 50% ORM, three levels of ability of *Ss*, and learning to one or three successive errorless repetitions. Seven nonsense syllable-adjective pairs were learned. None of the variables had significant effects upon learning, recall, or relearning.

In Exp. II, 90 *Ss* were randomly assigned to 100%, 50%, and 25% ORM conditions and two order conditions, NS-Adjective, and Adjective-NS. Mean trials to mastery were an inverse function of %ORM, although the difference between 100% and 50% ORM was surprisingly small, as in Exp. I. Adjective-NS pairs were significantly harder to learn than the NS-Adjective pairs but ORM × S-R Order was not significant.

In Exp. III, 100%, 50%, and 25% ORM, 1:1, 2:2, and 4:4 rates of presentation, and 4, 6, and 8 nonsense-syllable pairs of low association value were studied in a factorial design with 108 *Ss*. Trials were given continuously with no intertrial interval. Mean trials to criterion were an inverse function of %ORM. Increases in rate of presentation and length of list increased mean trials to criterion, but rate of presentation failed to affect rate of learning when total elapsed time to criterion was used as the measure, suggesting that *Ss* assimilated the material at a constant rate which was independent of rate of presentation. The Rate × Length interaction was significant, with trials to criterion as the measure.

References

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

Feldman, S. M., & Underwood, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53,** 11–15.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–267.

Goss, A. E., Morgan, C. H., & Golin, S. J. Paired-associates learning as a function of percentage of occurrence of response members (reinforcement). *J. exp. Psychol.*, 1959, **57,** 96–104.

Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27,** 453–464.

Hovland, C. I. Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable presentations. *J. exp. Psychol.*, 1938, **23,** 172–190.

Hovland, C. I. Experimental studies in rote-learning theory: VII. Distribution of practice with varying lengths of list. *J. exp. Psychol.*, 1940, **27,** 271–284.

Hovland, C. I. Experimental studies in rote-learning theory: VIII. Distributed practice of paired-associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39,** 714–718.

Kimble, G. A., & Dufort, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50,** 361–368.

Noble, C. E., & Stockwell, F. E. Stimulus vs. response meaningfulness in paired associate verbal learning. *Amer. Psychologist*, 1957, **12,** 425. (Abstract)

Schulz, R. W., & Runquist, W. N. Learning and retention of paired adjectives as a function of percentage occurrence of response members. *J. exp. Psychol.*, 1960, **59,** 409–413.

Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.

Trapp, E. P., & Kausler, D. H. A revision of Hull's table of associative values for 320 selected nonsense-syllables. *Amer. J. Psychol.*, 1959, **72,** 423–428.

Weaver, W. The mathematics of communication. *Scient. Amer.*, 1949, **11,** 180–184.

(Early publication received August 19, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 290–294

THE EFFECT OF CS PRESENCE DURING DELAY OF REWARD ON THE SPEED OF AN INSTRUMENTAL RESPONSE [1]

MORTON RIEBER [2]

Iowa Child Welfare Research Station, State University of Iowa

The effects of delay of reward on response strength has received considerable attention in S-R learning theory. Most of the investigations of this variable in the instrumental conditioning situation have employed animals as *S*s. The results of such studies (Harker, 1956; Perin, 1943; Ramond, 1954) indicate that, in general, delay of reward has a detrimental effect on response strength. The effects of delay on response strength in human *S*s is not as clear-cut. Lipsitt, Castaneda, and Kemble (1959) reported that children responded significantly more slowly to a light associated with delay than to one associated with immediate reward. In two similar experiments (Lipsitt, 1957; Lipsitt & Castaneda, 1958) no significant difference was found in the speed of response to the two lights, but when the children were given a choice as to which stimulus was to be presented (Lipsitt & Castaneda, 1958) the immediately reinforced one was chosen significantly more frequently.

It has been suggested (Spence, 1956) that a major effect of delay is to facilitate the establishment of competing responses to goalbox cues. It is hypothesized that an animal experiencing a delay of reward will make various responses in the presence of goalbox stimuli during the delay period. Assuming that such responses become associated with the goalbox cues, and generalize to the runway cues, they should interfere with the appropriate running response and slow it down. Similar effects might be expected for children in a delay of reward situation if the cues which elicit the appropriate response (CR) are also present during the delay interval. That is, the presence of such cues during delay would permit the association of inappropriate, competing responses to them. Hence, it would be expected that interference with the CR would be an increasing function of the similarity between the cues present during the delay period and those which elicit the CR.

The present experiment was designed to test the hypothesis that the effect of delay of reward in children is to facilitate the association of competing responses with the stimuli which elicit the CR. A simple instrumental conditioning situation was used in which *S*s were required to make a motor response to the onset of a CS. There were three experimental groups. Group DS experienced a delay of reward and the CS remained on during the delay period. Group D also experienced a delay of reward, but the CS terminated with the response and was not present during the delay period. Group IM was reinforced immedi-

[1] The present paper is a portion of a dissertation submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the PhD degree in the Iowa Child Welfare Research Station. The writer expresses his sincere appreciation to Charles C. Spiker who directed the dissertation.

[2] Now at Houston State Psychiatric Institute.

ately following the response and the CS terminated with the response. Speed of response was used as the criterion measure.

The effect of changing the stimulus conditions with the onset of the delay period may be evaluated by comparing the performances of Group D and DS. Any competing responses made by Group DS during the delay period may become conditioned to the CS. Group D could also have made competing responses during this interval but the CS was not present and the conditioning of these responses to it would have been minimal. Thus, to the extent that Group DS responds more slowly than Group D, there is support for the hypothesis that one of the major effects of delay of reward is to condition to the CS responses which compete with the CR.

A comparison of the performance of Group D with that of Group IM provides an indication of the effectiveness of delay when the association of competing responses is minimized.

METHOD

Subjects.—The *S*s were 89 kindergarten children attending the Cleveland Elementary School in Cedar Rapids, Iowa.[3]

Apparatus.—The apparatus consisted of a plywood box, approximately 20 × 20 × 20. A lever, located on the right side of the front face, protruded 8 in. On top of the lever there was a sponge rubber ball which provided a gripping surface. The lever could be moved downward a total distance of 15 in. Its complete depression activated a Hunter Decade interval timer which controlled a reinforcement dispensing device, and offset of a CS. The reinforcement dispenser was contained in the box and, when activated, it delivered a trinket into a container (goalbox) located at the bottom center of the front face. A 20-w. "black light" fluorescent tube was mounted on the top front edge of the box and wired so that it lit immediately. A reflector directed the light onto the goalbox and the front of the apparatus, both of which were coated with fluorescent paint which appeared light grey under ordinary light and bright green when illuminated by the "black light" tube. Much of the light emitted by the tube was visible and onset of this light served as the discriminative stimulus.

The apparatus allowed for automatic control of the interval between complete depression of the lever and delivery of the reward. Provision was also made for automatic control of the interval between complete depression of the lever and termination of the CS. Two Standard Electric clocks were wired into the circuit. The first clock started as soon as the stimulus light was turned on by *E*. When *S* started to depress the lever (move it ¼ in.), the first clock stopped and the second clock started. The second clock was stopped as soon as the lever was depressed through its full 15-in. sweep. Time recorded by the first clock will be referred to as starting latency and time recorded by the second clock will be referred to as movement time.

Procedure.—The 89 *S*s were randomly divided into three groups. The *S*s in Group D ($N = 30$) experienced a 12-sec. period of delay between their response to the CS and delivery of the reward (trinket). For this group the CS terminated simultaneously with the making of the response. Group DS ($N = 30$) also experienced a 12-sec. delay of reward, but the CS was present during the delay interval and terminated with delivery of the reward. Group IM ($N = 29$) experienced immediate reward and the CS terminated simultaneously with the delivery of the trinket. All *S*s were given 16 reinforced CS presentations (trials). The interval between CS onsets was kept constant for all groups and was approximately 20 sec. in duration. Starting and movement times were recorded for each *S* on each trial. Twelve of these measures were not recorded because of *E*'s error or apparatus failure.

The *S*s were brought into a semidarkened experimental room one at a time. The *E* told *S* that when the CS was turned on he was to press the lever down once, and this response would be followed by ejection of a trinket into the goalbox. The *S* then received one nonreinforced demonstration, without presence of the CS, in which he was asked

[3] The writer is grateful to H. C. Reid, Superintendent of Schools, Cedar Rapids, Iowa, for his permission to use the children as *S*s in this experiment. Appreciation is also expressed to Catherine Eckstein of the Cleveland Elementary School, and to her staff, for their fine cooperation and assistance.

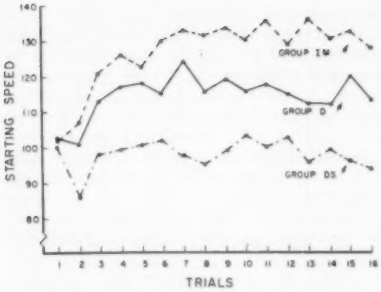

FIG. 1. Mean starting speeds per trial for each experimental condition.

to "try it" by pushing the lever all the way down. The *E* then reminded *S* that if he pressed the lever when the CS was on, a trinket would be ejected which he could keep. The *S* was also asked to keep his hand on his lap when the CS was not on. The *E* said "O.K." approximately 2 sec. before the onset of the CS, for each trial. This statement served as a ready signal.

RESULTS

The data were transformed by taking the reciprocal of the starting latency and movement time measures for each *S* on each trial and multiplying it by 100, yielding starting speeds and movement speeds. This transformation was carried out in order to better approximate a normal distribution. In cases where a score was missing, the trial mean for the group was used in the analysis of the data and the appropriate degrees of freedom were deducted from the error (w).

Starting speed.—Figure 1 presents the mean starting speeds for the three groups as a function of trials. The major interest was in the comparisons of Group D with Group DS, and of Group D with Group IM, on each of the 16 trials. This necessitated the use of two separate analyses of variance rather than an overall analysis. The first of these analyses (Table 1) indicated that Group DS responded significantly more slowly than Group D ($P < .001$). The Groups × Trials interaction proved to be nonsignificant ($F = 1.29$, $df = 15/867$). The Trials effect proved to be significant ($P < .001$).

The results of the second analysis (Table 1) indicated that Group D responded significantly more slowly than Group IM ($P < .05$). The Groups × Trials interaction was also significant ($P < .05$), as was the effect of Trials ($P < .001$).

A *t* test of the difference between the mean of the first two and last two trials was made for each of the groups separately. The results indi-

TABLE 1

ANALYSES OF VARIANCE COMPARING GROUPS D VS. DS AND D VS. IM ON STARTING SPEEDS OVER TRIALS

Source	Group D vs. DS			Group D vs. IM		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *Ss*	59			58		
Groups (G)	1	67,218	12.02***	1	34,845	4.34*
Error (b)	58	5,590		57	8,026	
Within *Ss*	897			883		
Trials (T)	15	1,217	3.64***	15	3,300	9.48***
G × T	15	430	1.29	15	579	1.67*
Error (w)	867	334		853	347	
Total	956			941		

* $P < .05$.
*** $P < .001$.

cated that both Group D and Group IM responded significantly faster on the last two trials ($P < .001$). There was, however, no significant difference between the means for the first and last two trials for Group DS ($t = 1.10$, $df = 29$).

Movement speed.—Figure 2 presents the mean movement speeds for the three groups as a function of trials. These data were analyzed in the same manner as the starting speed data. Table 2 summarizes the comparison of Group D with Group DS on each of the 16 trials. The analysis indicated that Group DS responded significantly more slowly than Group D ($P < .005$). The Trials effect was also significant ($P < .001$), as was the Groups × Trials interaction ($P < .005$).

A comparison of Group D with Group IM (Table 2) failed to show a significant difference between them ($F < 1.0$). The Groups × Trials interaction also proved nonsignificant ($F = 1.48$, $df = 15/855$), but the Trials effect was significant ($P < .001$).

Separate t tests comparing the means of the first and last two trials indicated that all three groups responded significantly faster on the last two trials ($P < .001$).

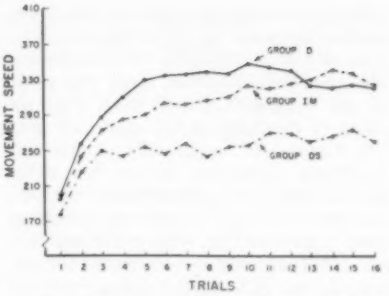

FIG. 2. Mean movement speeds per trial for each experimental condition.

DISCUSSION

The hypothesis that delay of reward results in the conditioning of competing responses to the CS leads to the prediction that Group D would respond faster than Group DS. The results of the present study confirm this prediction with respect to both starting speed (Fig. 1) and movement speed (Fig. 2).

The relative performances of Groups D and IM may be explained in terms of response competition, provided two further assumptions are made: (*a*) that competing responses are more likely to develop for Group D than for Group IM; and, (*b*) that competition is likely to have a greater effect on CR evocation (as measured by starting speed) than on CR duration (as measured by movement speed). The first assumption becomes plausible if it is noted that the stimulus

TABLE 2

ANALYSES OF VARIANCE COMPARING GROUPS D VS. DS AND D VS. IM ON MOVEMENT SPEEDS OVER TRIALS

Source	Group D vs. DS			Group D vs. IM		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *Ss*	59			58		
Groups (G)	1	1,020,90	9.51**	1	48,898	0.43
Error (b)	58	107,431		57	111,833	
Within *Ss*	893			881		
Trials (T)	15	53,346	17.76***	15	82,494	26.39***
G × T	15	7,305	2.43**	15	4,620	1.48
Error (w)	863	3,004		851	3,126	
Total	952			939		

** $P < .005$.
*** $P < .001$.

present immediately prior to the lever-pulling response includes the apparatus cues as well as the light. If we represent the stimulus situation after the onset of the light as CS, and that after the offset of the light as CS′, and note that for Group D potentially interfering responses occurring during the delay are followed by reinforcement, then we should expect such responses to become conditioned to CS′. Since CS and CS′ are relatively similar, it would be expected that CS should elicit these responses through stimulus generalization. The same analysis holds for Group IM except that competing responses (made during the intertrial interval) are not followed by reinforcement and should not be strongly conditioned to CS′.

The second assumption is based on an interpretation by Spence (1956, pp. 116ff.) which maintains that starting speeds are inversely related to the number and relative strengths of competing responses aroused by the CS, but that once the CR is initiated, the probability of occurrence of competing responses diminishes and remains at a relatively low level until the ongoing response is terminated. This assumption is also consistent with the fact that Group DS, which should have developed the greatest amount of competition for the CR, shows a significant increase over trials in movement speed, but no significant increase in starting speed.

Given these two assumptions, the explanation of the results for Group IM and D is straightforward. Group IM has a faster average starting speed than Group D because of the lesser amount of competition for Group IM than for Group D. There is no difference between the two groups in movement speed because competing responses do not affect movement speed as much as they affect starting speeds. Finally, the slower average movement speed of Group DS, as compared to either Group IM or Group D, is attributed to the greater amount of competition which develops for this group.

Summary

Three groups of kindergarten children were required to press a lever to the onset of a CS. Delivery of a trinket served as a reinforcer. Group DS experienced a delay of reward and the CS remained on during the delay interval. Group D also experienced a delay of reward, but the CS terminated with the response and was not present during the delay interval. Group IM was reinforced immediately following the response and the CS terminated with the response.

Two measures of response strength were recorded. Starting speed was a measure of the interval between onset of the CS and initiation of the response. Movement speed was a measure of the interval between initiation and completion of the response.

Results with the starting speed measure indicated that Group IM responded significantly faster than Group D, and that Group D responded significantly faster than Group DS.

A comparison of Group D with Group IM on movement speed indicated that there was no significant difference between these two groups. However, Group D had a significantly faster movement speed than Group DS.

It was suggested that the major effect of delay of reward, in this situation, was to facilitate the association of competing responses with the CS.

References

Harker, G. S. Delay of reward and performance of an instrumental response. *J. exp. Psychol.*, 1956, **51**, 303–310.

Lipsitt, L. P. Effects of differential (delayed reward) instrumental training on subsequent performance of children in a discrimination task. Unpublished doctoral dissertation, State University of Iowa, 1957.

Lipsitt, L. P., & Castaneda, A. Effects of delayed reward on choice behavior and response speeds in children. *J. comp. physiol. Psychol.*, 1958, **51**, 65–67.

Lipsitt, L. P., Castaneda, A., & Kemble, J. D. Effects of delayed reward pretraining on discrimination learning in children. *Child Develpm.*, 1959, **30**, 273–278

Perin, C. T. The effect of delayed reinforcement upon the differentiation of bar responses in white rats. *J. exp. Psychol.*, 1943, **32**, 95–109.

Ramond, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, **47**, 248–251.

Spence, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

(Received February 8, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 295-299

INDUCED MUSCULAR TENSION, INCENTIVE, AND BLINK RATE IN A VERBAL LEARNING TASK

JOSEPH B. SIDOWSKI AND CONRAD NUTHMANN

San Diego State College *Philco Corporation*

Several theoretical attempts have been made to explain the facilitating effects of induced muscular tension upon performance in both motor and verbal tasks (Courts, 1942a, 1942b; Freeman, 1933). The most recent theoretical development has been Meyer's (1953) theory of response interaction which attempts to account for the effects of induced tension upon both learned and unlearned responses. In brief, the theory states that induced muscular tension affects the rate of acquisition by promoting response elicitation. The theory attributes the interaction of simultaneous responses to a convergence of impulse patterns upon the motor pathways of the central nervous system. Interference with performance is held to result from a facilitation of competing responses. Obviously, the effects of induced tension cannot be the same throughout practice, since if an input captures an optimal population of neurons, added excitation triggers only neurons which contribute to the production of errors. This would tend to explain the results of several studies (Courts, 1942a, 1942b; Freeman, 1933) which show tension to be a valuable asset during early stages of practice but a hindrance to skilled performance.

Meyer also extends his analysis to learning situations in which artificial tension is absent. In instances where we find factors such as prolonged work periods, distractions, increased incentives, etc., additional tension is assumed to contribute to response elicitation. Finally, the theory provides a rationale for the use of eyeblink rate as an index of generalized muscular tension (Lovaas, 1960; Meyer, Bahrick, & Fitts, 1953).

The present study was designed in order to evaluate the influence of induced muscular tension and incentive upon performance and blink rate in a rote serial learning task. Within the framework of Meyer's theory, the following predictions were made: (*a*) learning rates would be higher when induced tension or incentive were introduced, (*b*) a combination of tension plus incentive would produce higher rates than either alone, and (*c*) a serial-position curve would result under all conditions for eyeblinks with perhaps a more pronounced peaking in the tension plus incentive condition.

METHOD

Subjects.—Twenty-four introductory psychology students at the University of Wisconsin served as *Ss* and received class points for doing so. Each *S* participated in the experiment at the same hour for five consecutive days. On Day 1, the practice day, all *Ss* were familiarized with the procedure. During the next four days each *S* served under each of the four experimental conditions.

Task.—The main task for *S* was to learn a list of nonsense syllables while serving under each of the following conditions: Control (merely learn a list), Tension (induced muscular tension while learning), Incentive (reward for fastest learner), and Tension + Incentive (reward for fastest learner while using induced tension).

Apparatus and procedure.—The experimental room was arranged so that blinks could be counted while *S* anticipated nonsense syllables. An adjustable chair for *S* was placed in front of a table which supported

the task apparatus and an oval, roofed cubicle of white artboard. A headrest was provided for *S* on the back of the chair, and a Stoelting hand dynamometer was mounted on the right arm rest. A metal clamp attached to the dynamometer could be adjusted so that when *S* squeezed to the prescribed level an auditory signal was presented. The auditory signal, a buzzing relay, was presented as long as *S* maintained the pressure within 3 kgm. of the prescribed level. The first syllable of the list (anticipatory cue) was a signal for *S* to squeeze and anticipate the syllables.

During the two tension conditions, *S* squeezed and kept the buzzer on during a trial. The hand was relaxed during the rest intervals between trials. During the Control and the Incentive conditions the buzzer was presented automatically by *E* throughout each trial in order to control for any effects that the buzzer may have had during the Tension conditions. The level of induced tension was determined by having *S* squeeze the hand dynamometer as hard as possible for 30 sec. One-fourth of the reading, taken at the end of the 30-sec. period, was used as *S*'s level of tension. A measure of *S*'s squeeze was taken each day prior to the beginning of the experimental period regardless of the condition for that day. In order to allow for dissipation of the tension effects, a period of 5 min. elapsed prior to beginning the experimental session for that particular day.

Five lists of nonsense syllables, one practice and four experimental, taken from Glaze (1928) and of high association value (87–100%), were used. A different consonant started each of the syllables in a list with the vowels being used about equally often. Each list consisted of 10 nonsense syllables. Syllables were presented auditorily by means of a Webcor tape recorder at a 4-sec. rate with an 8-sec. rest interval between trials. During the rest interval of the tension conditions, *S* relaxed his hand. Since the first syllable served as an anticipatory cue, *S* learned only 9 items. The *S*s learned each list until all syllables were correctly anticipated during a single trial. Since the syllables were spelled out to *S* from the tape recorder, all responses made by *S* were also spelled.

The interior of the *S*'s cubicle was illuminated by a 60-w. bulb attached to the roof. The *E* could observe *S*'s eyelids through a one-way vision screen covering the aperture cut into the near wall of the cubicle. The number of blinks was counted by *E* and automatically recorded outside of the room. A record, therefore, was available for the number of blinks made during the intersyllable intervals and during the rest intervals. At no time during or after the experiment was *S* given any indication that blinks were being counted. An eyeblink was defined as one complete closure of the eyelids. The above technique for measuring blinks was identical with the method used by Meyer, Bahrick, and Fitts (1953).

Design.—A single 4 × 4 greco-latin square design repeated six times was used. Latin letters represented the four experimental conditions while the Greek letters represented the four lists of syllables. The *S*s were told to learn the list as quickly as possible under all conditions. During the tension conditions, however, *S*s were also told to squeeze the hand dynamometer and to keep the buzzer on throughout a trial. During the Control condition, *S*s were merely told to learn the list. Incentive instructions were identical with the Control except that *S*s were also told that the fastest learner would receive an additional class point. During the Tension + Incentive condition *S*s squeezed the dynamometer and were offered the same reward for fast learning as in the Incentive condition. The two incentive conditions, therefore, afforded each *S* the opportunity of receiving $2.00 and two extra class points. At the conclusion of the sessions, all *S*s were questioned concerning the value of the money and class points in order to determine the incentive value of the rewards for each *S*. All *S*s verbally expressed interest in attempting to win the money and points.

Two *E*s were present throughout the experiment. One *E* counted blinks while the second *E* recorded the number of correct anticipations on the response sheets. Neither *E* was visible to *S* during the learning sessions.

Results and Discussion

Learning data.—Means and *SD*s for errors and trials to mastery are presented in Table 1. Analysis of variance for the total number of errors (presented in Table 2) showed significant Conditions and Day effects, with the overall number of errors significantly decreasing over the four experimental days. Subsequent tests for differences between ranked means (Duncan, 1955) showed the Tension condition producing significantly more errors than the Control and Tension

+ Incentive conditions ($P < .05$). Similar statistical analyses on the number of errors per trial per *S* showed Tension producing significantly more errors than all other conditions.

Since it was possible that tension might facilitate performance early in the learning process and inhibit performance later, trials to reach successive criteria of 3, 6, and 9 correct were also subjected to analyses. Although the data indicated a consistent superiority for the Control and Tension + Incentive conditions, analyses of variance showed: (*a*) no significant differences in number of trials to reach the criterion of 3 correct, (*b*) a significant difference in the number of trials to reach 6 correct ($F = 8.35$, $df = 3/60$, $P < .01$), and (*c*) the differences in number of trials to reach the criterion of 9 correct fell just short of the 5% level of significance (this analysis, however, showed a significant decrease in the total number of trials to learn over the four experimental days, $P < .01$). Subsequent tests (Duncan, 1955) on trials to reach 6 correct showed: (*a*) the Tension condition required significantly more trials than all other conditions ($P < .01$), (*b*) the Incentive condition required more trials than Control ($P < .05$), and (*c*) there were no other statistically significant differences.

TABLE 1

MEANS AND *SD*s OF TOTAL ANTICIPATION ERRORS AND TRIALS TO LEARN FOR EACH CONDITION

Conditions	Errors		Trials	
	Mean	*SD*	Mean	*SD*
Control	29.42	15.70	8.63	3.22
Tension	41.21	20.15	10.04	4.26
Incentive	35.75	17.23	9.67	3.50
Tens. + Inc.	33.50	21.06	9.25	4.44

TABLE 2

ANALYSIS OF VARIANCE FOR TOTAL ANTICIPATION ERRORS

Source	*df*	*MS*	*F*
Order of presentation	3	966.00	1.25
Between *Ss* in same order	20	781.75	
Days	3	854.33	5.40**
Syllable lists	3	371.00	2.35
Conditions	3	580.33	3.67*
Square uniqueness	3	307.67	1.94
Residual	60	158.20	
Total	95		

* $P < .05$.
** $P < .01$.

Eyeblink data.—The number of blinks per trial per *S*s (summed over the 24 *S*s) for each serial position are presented in Fig. 1. The expected bowed serial curves with more pronounced peaking for the tension conditions did not materialize (the related error data showed the classical bowed serial effect with peaking at the sixth syllable for all groups). The means and *SD*s for blinks per *S* per trial are shown in Table 3 for

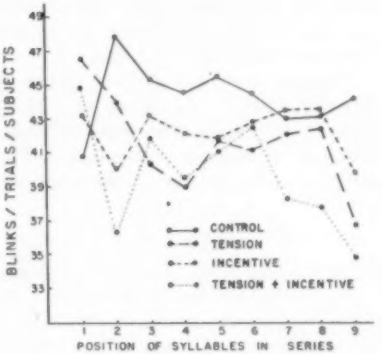

FIG. 1. The number of eyeblinks per trial per *S*s are presented for each serial position. Each of the points plotted above resulted from summing the individual blinks per trial values over the 24 *S*s at each serial position.

each condition. Analyses of variance were carried out on the blinks per *S* per trial measures (disregarding serial position) and on the total number of blinks. The results of the analyses indicated no significant differences between conditions for either measure. Analysis of total number of blinks, however, showed a significant decrease in blinks over the four experimental days ($F = 6.06, df = 3/60$, $P < .01$). Similar analyses of blinks to reach criteria of 3 and 6 correct showed no differences which could not be credited to total time to reach criterion.

In order to determine if blinks increased or decreased in frequency during learning, blink data for the first and last learning trials were summed for each *S*. The means for each condition are presented in Table 4. An analysis of variance for repeated measures showed a significant decrease in blinks from the first to the last learning trials ($F = 7.02$, $df = 1/23$, $P < .05$). No other statistically significant differences were found. An additional analysis for repeated measures was carried out over the first four trials and showed no significant differences in blinks between conditions or trials.

Several previous studies (King & Michels, 1957; Meyer et al., 1953) showed significant differences in blinks counted during rest intervals. The number of blinks during the 8-sec. intertrial rest intervals, therefore, was also subjected to analyses in the present study. No statistically significant differences were found between any of the conditions. (In the cited studies *S*s were required to perform visual tasks during the periods preceding the rest intervals.)

TABLE 3

EYEBLINKS PER *S* PER TRIAL (OVER ENTIRE TRIAL) FOR EACH CONDITION

Conditions	Mean	*SD*
Control	16.65	10.08
Tension	15.60	10.06
Incentive	15.85	10.10
Tens. + Inc.	14.91	9.86

TABLE 4

MEAN NUMBER OF BLINKS FOR FIRST AND LAST LEARNING TRIALS FOR EACH CONDITION

Conditions	First Trial	Last Trial
Control	18.12	15.50
Tension	16.83	14.13
Incentive	16.38	15.16
Tens. + Inc.	16.31	13.42

With regard to the blink data, one could say that the tension and incentive conditions were actually relaxing since they generally reduce blink rate (though not significantly so) and therefore should produce worse performance. This, however, is rather hard to believe. It is more likely that the superimposed tasks in the present study involve response patterns which do not include eyeblinks. (It should be noted, however, that blinks significantly decreased from the first to the last trials. Since trials to learn significantly decreased over days, the decrease in blinks over the same period can be attributed to total time to reach criterion.)

With regard to variations in verbal performance from condition to condition, it appears that grip tension consistently interfered with performance and that the incentive reliably interfered during part of the learning trials. This being the case, Tension + Incentive should interfere more than Tension alone. The results, however, indicate that this is not the case and that incentive and grip tension are not additive.

In conclusion, the results of the present study, fail to support the theory of simultaneous interaction of neural responses (Meyer, 1953).

Summary

Twenty-four college students, serving as their own controls, learned lists of nonsense syllables under each of the following conditions: (*a*) induced muscular tension, (*b*) monetary incentive, (*c*) combined tension plus incentive, and (*d*) no special incentive or induced tension (Control). Eyeblinks were counted during the learning trials and intertrial rest intervals.

The learning data showed: (*a*) the induced muscular tension consistently interfering with learning, (*b*) incentive reliably interfering during part of the learning sessions, and (*c*) no significant difference between the Control and Tension + Incentive conditions. The blink analysis showed no statistically significant differences between conditions either during learning or during the rest intervals. An expected bowed serial curve for blinks, paralleling the curves for the rote serial learning data, did not materialize.

References

Courts, F. A. The influence of practice on the dynamogenetic effect of muscular tension. *J. exp. Psychol.*, 1942, **40**, 504–511. (a)

Courts, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347–367. (b)

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

Freeman, G. F. The facilitative and inhibitory effects of muscular tension upon performance. *Amer. J. Psychol.*, 1933, **45**, 17–52.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

King, D. C., & Michels, K. M. Muscular tension and the human blink rate. *J. exp. Psychol.*, 1957, **53**, 113–116.

Lovaas, O. I. The relationship of induced muscular tension, tension level, and manifest anxiety in learning. *J. exp. Psychol.*, 1960, **59**, 145–152.

Meyer, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.

Meyer, D. R., Bahrick, H. P., & Fitts, P. M. Incentive, anxiety, and human blink rate. *J. exp. Psychol.*, 1953, **45**, 183–187.

(Received February 12, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 300-303

DENOTATIVE MEANING ESTABLISHED BY CLASSICAL CONDITIONING[1]

ARTHUR W. STAATS, CAROLYN K. STAATS, AND WILLIAM G. HEARD

Arizona State University

Previous studies of Staats and Staats (1957, 1959) have shown that the connotative factors of meaning established by Osgood and Suci (1955) may be considered implicit responses which can be classically conditioned. Osgood, Suci, and Tannenbaum (1957), however, state that verbal responses cannot always be differentiated in terms of their connotative meaning factors. For example, semantic differential profiles cannot differentiate satisfactorily among words like "hero," "success," "nurse," and "sincere." It might be stated that the inadequacy of connotative meaning in accounting for differences among words indicates that there are many other meaning responses involved which the semantic differential factors do not index. In order to determine if there are other meaning responses which follow the same principles as connotative meaning, it is important to see if nonaffective (referential) meaning components can also be classically conditioned. The present study tests the hypothesis that the denotative meaning component of angularity-roundness can be conditioned.

METHOD

Subjects.—There were 81 *Ss*, 47 from elementary psychology classes, and 34 from an education course.

Material.—The 34 *Ss* were used in constructing appropriate verbal materials for use in the experiment proper. The following procedure was designed to obtain a group of UCS words which was high in "angular" meaning and another group of words which was high in the opposite meaning of "round." At the same time, these two groups of UCS words were required not to differ on the three connotative factors of meaning: pleasant-unpleasant, active-passive, and weak-strong. This control was necessary to prevent the conditioning of connotative meaning in the procedure.

Ninety words, each paired with each of the four scales, were presented in a booklet for *Ss* to rate. Twenty-four "angular" and 24 "round" words were randomly distributed among 66 filler words and nonsense syllables. The sheets in the booklets were also randomized. The 12 "angular" words selected for the experiment proper were: SQUARE, BOX, ROOF, TRIANGLE, STEEPLE, DIAMOND, WINDOW, HALLWAY, ZIGZAG, BOOK, PYRAMID, WEDGE. The 12 "round" words were COIL, GLOBE, HUB, BARREL, BULB, TARGET, WHEEL, MARBLES, KNOB, HOOP, PEARL, BALL. The meaning scores for both groups of words on each of the four scales is given in Table 1. The table also includes the difference between the corresponding means for the two groups of words. Thus, a group of "angular" meaning words and a group of "round" meaning words were selected which did not differ to any great extent on any of the other scales.

Procedure.—The general procedure employed, using the 47 elementary psychology students, was the same as in the study of Staats, Staats, Heard, and Nims (1959)—except that *Ss* were run in groups instead of individually.

Four nonsense syllables were individually presented on a screen, with exposures of 5 sec. These syllables were the neutral conditioned stimuli: XEH, GIW, LAJ, and QUG. Approximately 1 sec. after the syllable appeared on the screen *E* pronounced a word aloud and *Ss* repeated the word aloud immediately. The *Ss* were told that they could learn each visually presented syllable by just looking

[1] This article arose from research sponsored by the Office of Naval Research under Contract Nonr-2794(02) and by the National Institute of Mental Health under Contract M-2381. The authors wish to thank Judson R. Finley for help in part of the data collection and statistical analysis.

TABLE 1

MEAN MEANING SCORES ON THE FOUR SCALES FOR THE TWO TYPES OF UCS WORDS

Type of UCS word	Semantic Differential Scales			
	Angular-Round	Pleasant-Unpleasant	Weak-Strong	Active-Passive
"angular"	2.19	3.27	5.15	3.56
"round"	5.91	3.18	5.02	3.04
Difference	3.72	0.09	0.13	0.52

at it, and that they could learn each word by hearing it and pronouncing it. The *Ss* were instructed not to make sentences out of the words or to associate them in any way.

Each of the syllables was given 12 conditioning trials. Order of the syllables was randomized except that no syllable appeared more than twice in succession. On each presentation of a given CS syllable it was paired with a different UCS word. Thus, 48 different words were used. The CS-syllables LAJ and GIW were always paired with UCS words having either "angular" or "round" meaning; XEH was always paired with words having "transportation" meaning: TRACTOR, CARRIAGE, WAGON, CART, CANOE, TRAILER, TAXI, BUGGY, BUS, JEEP, BARGE, SKATES; and QUG was always paired with words having "building material" meaning: PIPE, LATH, WOOD, PLASTER, ADOBE, WIRE, BOARDS, SCREWS, PUTTY, TILE, HINGES, BRICKS. For Group I (21 *Ss*) GIW was paired with UCS words of "angular" meaning and LAJ with the UCS words of "round" meaning. For Group II (26 *Ss*) this was reversed, GIW being paired with "round" words and LAJ with "angular" words. XEH and QUG were paired with the same set of words in both groups.

When the conditioning phase was completed, *Ss* rated the meaning of each of the four syllables on four semantic differential scales: angular-round, active-passive, weak-strong, and pleasant-unpleasant. The latter three scales represented the three connotative factors found by Osgood and Suci (1955). Each nonsense syllable was paired with each of the scales, thus 16 semantic differential ratings were obtained. The order of presentation of these syllable-scale items was randomized.

RESULTS

Five *Ss* from Group II were randomly eliminated to maintain equal groups in the counterbalanced design. The resulting N was 42. The means and *SD*s for each scale are listed in Table 2. The latin square design involved three variables for each meaning scale: Syllables (LAJ, GIW), Conditioned Meaning (angular-round, weak-strong, active-passive, or pleasant-unpleasant) and Groups (I, II). The analysis of variance (Lindquist, 1953, p. 278) for the data for each of the four scales revealed the following: (*a*) Meaning was conditioned as measured by the angular-round scale ($F = 10.45, df = 1/40\ P < .01$). The differences between nonsense syllables and groups were not significant. (*b*) Meaning was not conditioned significantly as measured by the pleasant-unpleasant scale ($F = 1.41$), nor were there significant syllable or group differences. (*c*) Meaning was not significantly conditioned as measured by the weak-strong scale ($F = 1.70$), nor were there significant syllable or group differences. (*d*) Meaning was not significantly conditioned as measured by the active-passive scale ($F = .58$), nor was

TABLE 2

MEANS AND *SD*s OF CONDITIONED MEANING SCORES

Scales	Group	Syllables			
		LAJ		GIW	
		Mean	*SD*	Mean	*SD*
Angular-Round	I	4.05	1.84	2.67	2.07
	II	2.90	1.67	4.48	1.87
Pleasant-Unpleasant	I	2.52	1.20	3.24	2.00
	II	3.29	1.60	3.24	1.70
Active-Passive	I	3.14	2.07	2.86	1.34
	II	3.95	2.10	4.24	1.84
Weak-Strong	I	4.05	1.92	5.05	2.00
	II	4.43	1.92	4.29	2.05

Note.—The Angular, Pleasant, Active, and Weak poles of the scales were scored 1.

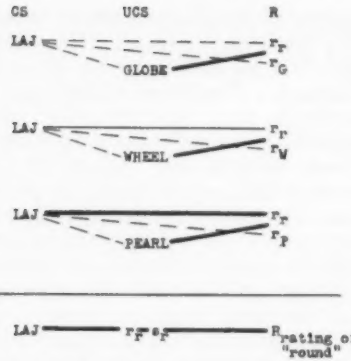

FIG. 1. The language conditioning of the "round" denotative meaning response. The nonsense syllable LAJ, in this example, is presented prior to the word GLOBE. GLOBE elicits a total meaning response. This is schematized in the figure as two component responses; the "round" meaning response component r_r, and the other distinctive responses that characterize the meaning of the word, r_G. The pairing of LAJ and GLOBE results in associations between LAJ and r_r, and LAJ and r_G. In the following presentations of LAJ and the words WHEEL and PEARL, the association between LAJ and r_r is further strengthened. This is not the case with associations r_G, r_W, and r_P, since they occur only once and are followed by other associations which are inhibitory. The direct associations indicated in the figure between the name and the individual words would also in this way be inhibited.

there a significant Syllable difference. The Groups variable was significant ($F = 5.80$, $df = 1/40$, $P = <.05$).

DISCUSSION

It may be concluded that what is commonly called denotative meaning may be conditioned to CS syllables to form their meanings. Moreover, the principles of denotative meaning conditioning appear to be the same as those involved in the conditioning of connotative meaning, i.e., classical conditioning. The way the conditioning was thought to have taken place in the experiment is schematized in Fig. 1.

A number of studies (e.g., Leuba, 1940; Lipton & Blanton, 1957; Phillips, 1958; and studies of sensory preconditioning) indicate that sensory processes may be conditioned and two of them (Lipton & Blanton, 1957; Phillips, 1958) indicate these responses may constitute the meaning of words. Mowrer (1960) has suggested that sensations may be conditioned and Staats (1959b) has suggested that one type of meaning response may be a conditioned sensory response. Thus, in terms of the present experiment, "angular" and "round" meaning may be considered to be implicit meaning responses of a sensory nature which have been conditioned to the words from past pairings of the words with stimulus objects having round or angular sensory qualities.

The present results also have some relevance for semantic differential measurements. First, the results indicate that although any one particular semantic differential scale may be measuring more than one meaning component, it is possible to preselect UCS words which will condition one component of meaning without conditioning others. Secondly, however, the results indicate that since specific meaning response components not among the connotative factors identified by Osgood et al. may be important in understanding language behavior, efforts to construct measures of word meaning based upon a small number of factors may be vitiated.

The results also indicate that the language conditioning of meaning was strong enough to resist from one to four extinction trials prior to measurement of the conditioned meaning; that is, each presentation of the syllable until the appropriate scale was reached must be considered an extinction trial.

As was discussed in an earlier article (Staats, Staats, & Heard, 1959) the results obtained in experiments using the method for language conditioning of meaning could be interpreted in word association terms, i.e., a set of UCS words as stimuli might elicit common word associates which might be conditioned to the syllable and mediate the semantic

differential scoring. This, however, seems a less likely alternative for the reasons cited in the earlier article and for those presented in fuller discussions of this interpretation (Staats, 1959a; Staats & Staats, 1959).

Summary

The general method was to pair a visually presented nonsense syllable as the CS once each with one of a set of 12 different auditorily presented UCS words having a common denotative component of meaning, i.e., either "round" or "angular." In the experiment each of two nonsense syllables was conditioned to each of the two components of meaning, to counterbalance the effect of the syllable. To test unequivocally the conditioning of denotative meaning the simultaneous conditioning of connotative meaning had to be eliminated, i.e., groups of UCS words were selected which varied on denotative meaning, but not on connotative meaning. Using this method of control, significant denotative meaning was conditioned as measured by semantic differential scales.

References

Leuba, C. Images as conditioned sensations. *J. exp. Psychol.*, 1940, **26**, 345–351.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Lipton, L., & Blanton, R. L. The semantic differential and mediated generalization as measures of meaning. *J. exp. Psychol.*, 1957, **54**, 431–437.

Mowrer, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

Osgood, C. E., & Suci, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

Phillips, L. W. Mediated verbal similarity as a determinant of the generalization of a conditioned GSR. *J. exp. Psychol.*, 1958, **55**, 56–62.

Staats, A. W. Meaning and word associations: Separate processes. Paper presented at the meetings of the American Psychological Association, Cincinnati, 1959. (a)

Staats, A. W. Verbal habit families, concepts, and the operant conditioning of word classes. ONR Technical Report, 1959, Arizona State University, Contract Nonr-2794(02). (b)

Staats, A. W., & Staats, C. K. Meaning and *m*: Correlated but separate. *Psychol. Rev.*, 1959, **66**, 136–144.

Staats, C. K., & Staats, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54**, 74–80.

Staats, A. W., Staats, C. K., & Heard, W. G. Language conditioning of meaning to meaning using a semantic generalization paradigm. *J. exp. Psychol.*, 1959, **57**, 187–192.

Staats, A. W., Staats, C. K., Heard, W. G., & Nims, L. P. Replication Report: Meaning established by classical conditioning. *J. exp. Psychol.*, 1959, **57**, 64.

(Received February 12, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 304–308

RESISTANCE TO EXTINCTION AS A FUNCTION OF THE FIXED RATIO[1]

JOHN J. BOREN[2]

Columbia University

In 1938, Skinner described a schedule of reinforcement which he called fixed ratio. According to this schedule, reinforcement was delivered after a certain fixed number of responses. The term arose from bar-pressing experiments on another schedule, periodic or fixed interval reinforcement, in which the number of responses the animal made in the "extinction" period between reinforcements was observed. Skinner quantified these observations in terms of an extinction ratio: the ratio of extinction responses to reinforced responses. In other experiments, he controlled the extinction ratio by making the reinforcement contingent upon a fixed number of responses—thus a fixed (extinction) ratio schedule of reinforcement. The lever-pressing rate on this schedule accelerated abruptly after a brief postreinforcement pause and continued at a high level until the reinforcement. Animals trained on such a schedule emitted a large number of responses during extinction.

In the present experiment the number of responses to extinction following fixed ratio training was studied in a systematic manner. By conditioning groups of *S*s on different fixed ratio schedules and then extinguishing, the function relating the size of the fixed ratio to the resistance to extinction was determined.

In this paper the convention has been followed of defining the fixed ratio (FR) in terms of the extinction ratio. Thus, if every sixth response is reinforced, the FR is 5:1.

METHOD

Subjects and apparatus.—Thirty-six male albino rats from Carworth Farms were used as *S*s. At the beginning of experimentation *S*s were 130 to 150 days old.

The bar-pressing apparatus employed was slightly modified from that described by Frick, Schoenfeld, and Keller (1948) by enlarging the diameter of the bar to $\frac{1}{2}$ in. to prevent bar biting, and by requiring about 10 gm. force for bar depression. The food magazine was made to discharge a 1-grain pellet of Purina laboratory chow.

Six *S*s were tested simultaneously in a dimly illuminated experimental room. This room was sound resistant to external noises, and the hum of an electric fan plus the clicking of a dummy food magazine masked noises from within the room.

The FR reinforcement was programed automatically by relay circuitry outside the experimental room. A conditioned reinforcer, the sharp click of the food magazine solenoid, followed the reinforced bar depression immediately.

Procedure.—The 36 *S*s were divided randomly into six equal groups and placed in the individual living-experimental cages which they occupied throughout the experiment. Two *S*s were eventually discarded, one from Group I (FR 0:1) because of apparatus malfunctioning and one from Group VI (FR 20:1) for failure to maintain responding under FR training.

The following preliminary procedure was common to all groups. Six days prior to experimentation *S*s were put on a feeding regimen under which they were fed ground Purina laboratory chow for 1 hr. daily. Water was continuously available. This feeding regimen was continued for the duration of the experiment. On Days 1 and 2 of experimentation the unconditioned rate of bar pressing (operant level) was measured for 30 min. each day. On these and all fol-

[1] This study was submitted as part of the research requirements for the PhD degree. I am indebted to Fred S. Keller for invaluable guidance during the course of this research.

[2] Now at Merck Institute for Therapeutic Research, West Point, Pennsylvania.

lowing days the experimental sessions began when *Ss* had been deprived of food 22 hr. No significant differences in operant level among groups were found. Following each operant level session and after the bar had been removed, eight pellets were discharged one at a time from the food magazine at approximately 2-min. intervals. This procedure trained *Ss* to approach the food tray and eat at the sound of the magazine click. On Day 3 all *Ss* were given 20 continuous reinforcements for bar pressing (i.e., FR = 0:1, where a food pellet was discharged after each bar press). With this session the preliminary training was concluded.

Group I then received 540 additional continuous reinforcements, 40 on Day 4 and 50 on each of 10 succeeding days. Groups II–VI first received 40 reinforcements at an FR of 2:1 (so that *Ss* to be shifted later to higher ratios would not extinguish) and were then placed on a final schedule for 500 more reinforcements. Group II continued on an an FR of 2:1, Group III was shifted to 5:1, Group IV to 9:1, Group V to 14:1, and Group VI to 20:1. Fifty reinforcements were delivered on each of 10 days. In summary, then, the groups received a total of 560 reinforcements but differed in how the last 500 were programed.

Following the above training procedure all groups were given five daily 1-hr. extinction sessions during which the food magazine did not operate.

Results

The response rates for the various groups at the end of the training period are shown in the lower curve of Fig. 1. These rates were averaged from the last two conditioning sessions after *Ss* were well trained on their ratio schedule. Figure 1 shows that as the size of FR was increased from group to group, the rate increased in a negatively accelerated fashion. The statistical significance of the functional relation was assessed by the nonparametric Jonckheere (1954) test for ordered sequence. The null hypothesis that the rates of the various groups were ordered at random with respect to FR was rejected at better than the .001 level. This result indicates that the size

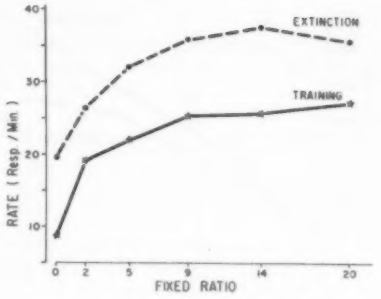

FIG. 1. The average response rate for each fixed ratio schedule. The solid curve represents the rate at the end of training while the dashed curve shows the rate for the first 2 min. of extinction.

of FR is an important determinant of the response rate.

In interpreting these data, two limitations must be kept in mind: (*a*) Figure 1 shows a *group* functional relation and does not necessarily represent the changes in rate which might be observed if an individual *S* were successively trained on the various ratio schedules. (*b*) the rates during training include the time to pick up the food pellet and perhaps to eat it (although rats may eat while pressing the bar). Since *Ss* in Group II (FR 2:1) were getting about 6 pellets/min while *Ss* in Group VI (FR 20:1) were receiving about 1 pellet/min, food-getting behavior occupied a greater proportion of time for the former group than for the latter. Thus, the rate at the end of training may be due in part to the proportion of time occupied by food-getting behavior rather than differences in rate while responding.

Since no pellets were delivered during extinction, the initial rate during the first extinction session provides a rate measure which, although it may be complicated by other factors, is free of food-getting behavior. The average rates for the

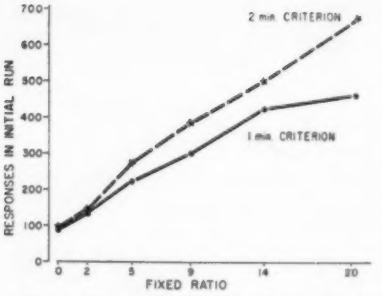

FIG. 2. The length of the initial run in the first extinction session for each fixed-ratio schedule.

various groups, calculated from the responses occurring in the first 2 min. of the first extinction session, are shown in the upper curve of Fig. 1. The Jonckheere test showed significance at better than the .001 level. As was true of the rate at the end of training, the rate at the beginning of extinction increased with the size of FR and the curve was negatively accelerated.

The individual extinction curves of Group I were typical of continuous reinforcement and require no comment. The curves of Groups II–VI began with a high steady rate of responding. Two-thirds or more of the responses made in the entire session typically were emitted during this initial run. Figure 2 shows how the length of the run varied with the size of FR during training. Since defining the end of a run is somewhat arbitrary, two criteria were used. The run was defined by counting the number of responses from the beginning of the session up to the point where a 1-min. *or* a 2-min. pause occurred. Figure 2 shows that with either criterion the length of the run increased systematically with the FR schedule ($P < .001$, Jonckheere test). A comparison of the curves for the two criteria suggests that at the low ratios the 1-min. pause occurred quite close to the 2-min. pause. At the higher ratios, however, many responses might occur after the 1-min. criterion was met but before the 2-min. criterion could be satisfied.

In evaluating the over-all resistance to extinction, the most relevant measure is provided by the total number of responses made during the extinction sessions. The group means and *SD*s of the responses made in each session are presented in Table 1. When the Jonckheere test was applied to the group differences in total extinction responding over five sessions, the deviation from a random order was found to be significant at the .001 level. Thus, it is not likely that the extinction function is due to chance fluctuation. As another indication of the reliability of the function, the rank order of the size of the FR correlates exactly with the rank order of the number of extinction responses on Days 1 and 2, and on Day 3 the single inversion is between adjacent groups. The orderliness on separate days suggests considerable reliability in the increasing form of the functional relation. By Days 4 and 5 of extinction, however, the differences among the groups had

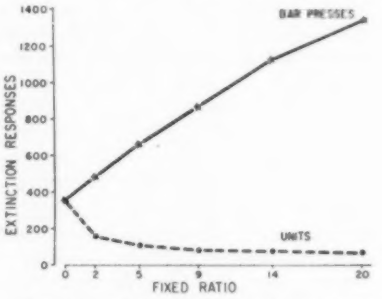

FIG. 3. Resistance to extinction as a function of the fixed ratio. In the upper curve are plotted the group means of the total number of bar presses made in the first three extinction sessions. In the lower curve are plotted the mean response units, as defined by Mowrer and Jones (1945).

TABLE 1

MEANS AND *SD*s OF RESPONSES MADE IN EACH DAILY EXTINCTION SESSION

Group	Day 1		Day 2		Day 3		Day 4		Day 5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
I (0:1)	162.8	66.3	109.2	41.3	84.8	40.4	93.8	54.9	85.2	41.5
II (2:1)	225.3	42.9	148.8	38.4	105.5	24.2	103.7	41.6	95.5	45.6
III (5:1)	396.2	96.5	164.0	57.2	111.5	43.3	117.5	44.4	67.8	24.6
IV (9:1)	453.3	34.9	218.3	34.0	189.5	33.0	156.7	48.8	128.8	42.0
V (14:1)	600.2	115.8	286.0	57.4	240.3	140.3	143.3	65.4	186.0	48.1
VI (20:1)	752.2	262.7	360.4	68.8	229.0	51.9	191.4	62.7	105.4	36.2

diminished, as might be expected after a large amount of extinction.

The total extinction responses for Days 1–3 are plotted against FR in the upper curve of Fig. 3. It is clear that the resistance to extinction increases with FR, at least over the range studied. The function is approximately linear but with a suggestion of negative acceleration. It might be pointed out that the largest FR used in this study (20:1) was carefully selected on the basis of pilot studies as a ratio which was small enough to maintain rapid responding in most *S*s. However, if the ratio were increased, a value could be reached which would not maintain a substantial rate (as shown by Ferster and Skinner [1957] and by my own pilot studies on two rats using a FR of 50:1) and which, therefore, would not produce a large extinction curve. Thus, considered over a much wider range of fixed ratios than was covered in the present experiment, it is likely that the extinction function reaches a maximum at a ratio somewhat greater than 20:1 and then declines as the ratio is increased to very high values.

DISCUSSION

The resistance to extinction data (Fig. 3 and Table 1) may be compared with Wilson's data (1954) for various fixed-interval schedules. The comparison can be made rather closely since his experiment and the present one were carried out in the same laboratory under very similar conditions. With fixed-interval reinforcement, the resistance to extinction rises sharply to a maximum at approximately 1 min. and then declines as the interval is lengthened. In contrast, the fixed ratio-extinction function (from FR 0:1 to 20:1) rises gradually in a linear fashion. Although the fixed ratio function will also pass through a maximum if the ratio is made very large, the two experiments show, over a rather wide range of values, quantitatively different relationships.

Resistance to extinction is sometimes held to depend on the amount of generalization between training conditions and extinction conditions (Mowrer & Jones, 1945). The closer the resemblance between conditioning and extinction, the greater is the persistence of the response in the face of nonreinforcement. When this theory (the generalization decrement hypothesis) is applied to a fixed-ratio schedule of reinforcement, the events which could generalize to extinction might be the number of responses without a reinforcement ("count"), or perhaps the maintenance of a high rate for a certain time (rate plus time). On the basis of either event, the generalization decrement hypothesis would predict that the higher the fixed ratio, the greater the resistance to extinction. This prediction is in accord with the data of the present experiment. On the other hand, the theory is vague enough to cover a wide variety of rising functions: a linear curve, a negatively accelerated curve, an S-shaped curve, etc. However, the theory would not be confirmed by a curve with a maximum.

Thus, it would fail to account for the relatively low resistance to extinction resulting from a large fixed interval, and from the analogous situation existing with a very large fixed ratio. It appears, then, that the generalization decrement hypothesis would require modification to account for the resistance to extinction generated by a wide range of fixed ratios or fixed intervals.

Mowrer and Jones (1945) proposed another theory of extinction based upon a consideration of the "response unit." According to their definition, a response unit is the total behavior which precedes the reinforcement, and this unit should hold together during extinction. Thus, if *S* is reinforced after every sixth bar press, the unit of six should be considered the "response." A group of *S*s reinforced for every sixth bar press, then, should make six times as many bar presses in extinction as a group reinforced for every bar press.

In order to evaluate this theory, a calculation was made upon the extinction data to yield the response unit curve shown in the lower part of Fig. 3. This curve was derived from the upper one by dividing the total number of extinction responses (bar presses) by the appropriate number of responses per reinforcement (the fixed ratio + 1). The Mowrer-Jones theory predicts a linear function with a slope of zero, but the empirical curve departs systematically from the prediction. Even at the larger fixed ratios, the curve decreases in an orderly way (although the change is difficult to see in the scale of this figure). Mowrer and Jones, who made a similar finding in their experiment, argued that another variable, the increased amount of effort in making the larger response units, might disturb predictions of the theory. Unfortunately, however, it is difficult to evaluate this argument since some studies show that the amount of effort is an important variable (Capehart, Viney, & Hulicka, 1958; Mowrer & Jones, 1943) while other studies do not (Applezweig, 1951; Maatsch, Adelman, & Denny, 1954). In addition, it is not clear whether a heavily weighted bar in these experiments is a good analogy to the fixed ratio schedule.

One other piece of evidence pertaining to the theory was gathered. The individual extinction records were carefully examined for bar-press groupings approximating the number of responses per reinforcement. No such groupings could be seen. Thus, the present experiment does not provide any clear evidence for the response unit hypothesis.

Summary

Groups of rats were trained in a bar-pressing apparatus on various fixed-ratio schedules of reinforcement ranging from continuous reinforcement to 20:1 and were then extinguished. The response rates, both at the end of training and at the beginning of extinction, increased with the fixed ratio. An approximately linear increasing relationship was found between the fixed ratio and the number of responses made in extinction.

References

Applezweig, M. H. Response potential as a function of effort. *J. comp. physiol. Psychol.*, 1951, **44**, 225–235.

Capehart, J., Viney, W., & Hulicka, I. M. The effect of effort upon extinction. *J. comp. physiol. Psychol.*, 1958, **51**, 505–507.

Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

Frick, F. C., Schoenfeld, W. N., & Keller, F. S. Apparatus designed for introductory psychology at Columbia College. *Amer. J. Psychol.*, 1948, **61**, 409–414.

Jonckheere, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

Maatsch, J. L., Adelman, M., & Denny, R. M. Effort and resistance to extinction of the bar-pressing response. *J. comp. physiol. Psychol.*, 1954, **47**, 47–50.

Mowrer, O. H., & Jones, H. M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33**, 369–386.

Mowrer, O. H., & Jones, H. M. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 293–311.

Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

Wilson, M. P. Periodic reinforcement interval and number of periodic reinforcements as parameters of response strength. *J. comp. physiol. Psychol.*, 1954, **47**, 51–56.

(Received February 17, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 309–314

DIFFERENTIAL RELATION OF LATENCY AND RESPONSE VIGOR TO STIMULUS SIMILARITY IN BRIGHTNESS DISCRIMINATION [1]

ALFRED CASTANEDA AND LEONARD WORELL

University of Texas *Oklahoma State University*

Several writers (Brown & Farber, 1951; Miller, 1944) have suggested that response competition (conflict) may provide a source of motivation in addition to those other motivational sources upon which the competing responses are based. On this basis, if an overt response has been initially aroused in competition with another, the motivational (energizing) effect of the conflict- or competition-produced motivational state might be expected to be reflected in some measure of its vigor. However, with one exception (Finger, 1941) most studies concerned with the simultaneous arousal of incompatible responses have reported a weakening of the overt response, e.g., lengthening of its latency, blocking, reversals in its direction, etc. The conditions of measurement may have been such, however, that they primarily reflected the inhibiting influence of the competing tendency on the overt response.

The study by Finger indicated that two measures (latency and force) of brightness discrimination performance in rats in the jumping-stand situation were differentially related to variations in similarity between the positive and negative cues. Following simple instrumental conditioning procedures with the positive stimulus, differences in similarity between the two cues were introduced during discrimination training. Latency of the jump response was shown to lengthen with increased similarity, a finding which is in accord with the assumption that the greater the generalization between cues, the greater the degree of competition between responses. The force of the jump response, on the other hand, was shown to *increase* with increased stimulus similarity. This finding is consistent with the hypothesis that the level of drive dependent on simultaneously aroused incompatible responses is inversely related to the magnitude of the difference in strength between them (Brown & Farber, 1951).

The present study is concerned with the hypothesis of competition-produced drive and extends selected aspects of Finger's procedure to motor conflict in human *Ss*. By means of simple motor responses *Ss* are required to discriminate between stimuli varying in brightness. Similarity between the stimuli is varied over two levels and the latency and vigor of the selective response are recorded. Since errors in discrimination are likely to be more frequent the greater the similarity between the stimuli, an attempt is made to minimize their incidence and influence. Such a consideration seems important on the possibility that one of the consequences of an incorrect response is to increase the vigor of subsequent responses. On the basis of the hypothesis that the level of motivation increases as the difference in strength between competing responses becomes smaller, increased similarity is expected to lead to a

[1] This research was done at the State University of Iowa.

longer latency, but a greater vigor of the selective response.

METHOD

Apparatus.—The apparatus consisted of two, 4-cm. diameter, circular milk-glass windows which were arranged 32 cm. horizontally from each other on a black panel and which, from a sitting position, were slightly below eye level. Illumination of the windows by single 7-w. incandescent lamps varying in brightness served as the stimuli. The handle of a Smedley-type hand dynamometer protruded approximately 13 cm. directly below each window and was the only of the dynamometer visible to *S*. A "start" platform, 18 × 18 cm., situated 13 cm. from either handle, resting on springs and a concealed microswitch, was fastened directly to the front of the panel. A ready signal, consisting of a 6-v. pinpoint of light was arranged centrally and above the stimulus windows.

A Standard Electric timer was energized simultaneously with the illumination of the windows and stopped when *S* removed his hand from the start platform. The lights in both windows were terminated by exertion of approximately 4 to 5 kg. pressure on either one of the two dynamometers. Grip pressures, in kg., could be read from the indicator needle which remained at the point of maximum pressure on the scale of the dynamometer. This arrangement for recording grip pressure is a modification of one described elsewhere (Castaneda, 1956). Latency, time from illumination of the windows to removal of the hand from the platform, was recorded to the nearest .01 sec. and converted to its reciprocal later. The ready signal preceded the onset of illumination by a 3-sec. interval.

The stimuli consisted of four different brightnesses which were obtained by means of appropriate settings of variable transformers in circuit with each lamp. The four different brightnesses, 2.8, 8.8, 90.9, and 291.4 ft-c, were obtained by measurements at the window with a MacBeth illuminometer. The two brightest stimuli (90.9 and 291.4) were separated by a difference of approximately .50 of a common log unit as were the two dimmest (2.8 and 8.8). The two extreme values (2.8 and 291.4) were separated by a difference of approximately 2.0 log units.

Procedure.—The experiment was conducted in a darkened room with the major source of light coming from the control units which provided sufficient light for reading instruments and recording. The *S*s were allowed approximately 3 to 5 min. dark-adaptation time during which instructions were read and the two practice trials described below were given. The *S* was informed that the experiment dealt with accuracy of visual discrimination. He was instructed to select the "brighter" or "dimmer" of two lights, depending on the condition to which he was assigned, by removing his hand from the start platform and gripping the handle of the dynamometer below the window containing the correct stimulus. The *S* was cautioned not to remove his hand from the platform until the location of the correct stimulus had been determined. The same (preferred) hand was used throughout the experiment. During the discrimination trials, reactions with latencies of less than .12 sec. were considered anticipatory errors. In these instances, which were fairly infrequent, *S* was again cautioned and the trial repeated. No mention was made of the fact that grip pressure was being recorded. The instructions carried no implication that speed was important. This aspect was considered necessary since preliminary work suggested that grip pressure depended on positioning of the hand on the handle which, in turn, appeared to be more variable under instructions implying speed.

The following aspects of the procedure were intended to minimize the incidence and influence of errors. Prior to the discrimination trials *S* was informed of the correct stimulus by use of the words "brighter" or "dimmer" in the instructions. Also, the correct stimulus was shown alone once in each window and *S* was required to make the appropriate response. Furthermore, during the discrimination trials, the lights in the window were always terminated regardless of whether or not the response was the correct one and *S* was never informed of the nature of his response. Thus, as far as this aspect of the procedure is concerned, the consequences of a correct or an incorrect discrimination were nondifferential.

During the discrimination trials the position of the correct stimulus was randomized under two predetermined orders with the restriction that the same stimulus appear not more than three times consecutively in the same position. A total of 48 trials was given and intertrial intervals were allowed to vary with the time required by *E* to read, record, and reset the apparatus for the next trial. Errors, when they occurred, were recorded for all *S*s.

Subjects and experimental design.—The *Ss* were 84 volunteer males from introductory psychology courses who were randomly assigned, 28 per group, to one of the following three groups. For Group I the stimuli consisted of the two extreme brightnesses, 2.8 and 291.4 ft-c. The stimuli for Group II consisted of the two lowest brightness values, 2.8 and 8.8, and for Group III, of the two highest, 90.9 and 291.4. Within each group the brighter of the two stimuli was designated the correct stimulus for half of the *Ss* and the dimmer for the remaining half. For convenience these subgroups will be identified by the letter B or D attached to the respective group number, e.g., I-B (two extreme brightnesses, brighter of the two correct). Within each of these six subgroups, half received one of the two orders of stimulus presentation by which the position of the correct stimulus was randomized and the remaining half the other.

RESULTS

The order of stimulus presentation will be ignored since it was not found to be related to either latency or grip pressure.

Latency.—Mean reciprocal latencies are presented separately for all subgroups in Fig. 1. in successive 8-trial blocks.

The *Fs* summarized from three separate analyses of variance (Lindquist, 1953, pp. 291–297) appear in Table 1. It can be seen that the effects of similarity (Groups), based on comparisons of Group I with Groups II and III separately, are highly reliable in each case. In view of the reliable higher-order interactions in these two comparisons the remaining significant lower-order interactions will be ignored. Inspection of the curves in Fig. 1 suggests that Groups I-B and I-D are the main contributors to the higher-order interaction. Not only is Group I superior to Groups II and III, but the difference between I-B and I-D also increases with successive blocks of trials. Hence, the interaction suggests that the stimulus intensity effect is dependent on the combined effects of similarity and stage of practice. In order to determine the degree of reliability of the interaction between the intensity of the correct stimulus and stage of training within each of the three major groups, analyses were performed on the data of Groups I, II and III separately. In the case of Group I the interaction was significant at the .001 level but it did not reach the .05 level in either Group II or III. Similarly, the effect of the intensity of the correct stimulus was significant at the .05 level only in the case of Group I.

In the Group II vs. Group III comparison, the only other significant

TABLE 1

Fs AND ERROR *MSs* FROM ANALYSES OF VARIANCE OF SPEED SCORES

Source	*df*	Groups Compared		
		I vs. II	I vs. III	II vs. III
Groups (G)	1	76.08***	65.93***	.74
Stimulus (S)	1	3.19	4.23*	.33
G × S	1	3.25	2.71	.49
Error (b) (*MS*)	52	*153.32*	*144.12*	*123.79*
Trials (T)	5	91.97***	110.26***	95.40***
T × G	5	4.10***	1.85	3.36**
T × S	5	2.80*	1.57	1.39
T × G × S	5	5.46***	5.64***	1.98
Error (w) (*MS*)	260	*4.04*	*4.24*	*3.36*

* $P = .05$.
** $P = .01$.
*** $P = .001$.

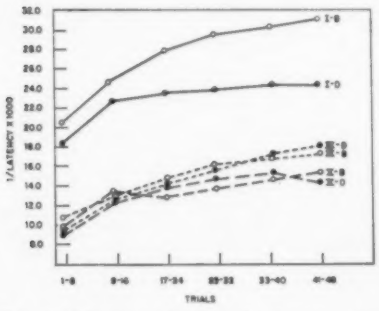

FIG. 1. Mean speed-of-reaction in successive 8-trial blocks.

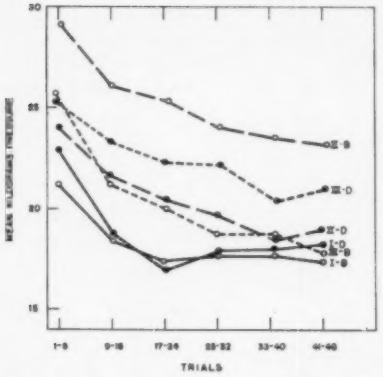

FIG. 2. Mean grip pressures in successive 8-trial blocks.

effect of interest is the Groups × Trials interaction. Inspection of Fig. 1 shows Group III to increase in its superiority over Group II with successive blocks of trials. Assuming the differences between the stimuli to be subjectively comparable in both groups, the effect suggests that stimuli selected from the more intense end of the brightness dimension are more likely to be associated with shorter latencies as training progresses. Finally, it may be noted that latencies become shorter with trials in all cases.

Grip pressure.—Mean grip pressure scores based on all 48 trials appear in successive 8-trial blocks in Fig. 2. In contrast to latency, pressure is positively related to stimulus similarity. Table 2 which summarizes the *F*s from the separate analyses of variance, indicates that the effects of similarity (Group I vs. II and I vs. III) are significant at the .01 and .05 levels, respectively. In all comparisons the effect of trials is highly significant and inspection of the curves shows pressures to decrease over trials in all cases. This effect may be attributable to fatigue and perhaps, for some *S*s, to learning that only slight pressure was required to terminate the lights.

The only other reliable effect, the Groups × Stimulus interaction in the comparison involving Groups II and III, was not entirely unexpected on the basis of preliminary work. Examination of the results in Fig. 2 indicates that while the brighter of the two stimuli (II-B) was associated with higher pressures in Group II, opposite effects appeared in Group III. Higher pressures were associated with the dimmer of the two stimuli (III-D). These differences were not significant when evaluated by *t* tests.

Although the procedure attempted to minimize the incidence of errors in discrimination, errors did occur. Furthermore, a higher frequency of errors occurred in the two groups which discriminated between the more similar stimuli. No errors occurred in Group I; 1 *S* made four errors while 4 *S*s made one error each in Group II; 1 *S* made 11 errors and 5 *S*s made one error in Group III. Although the procedure attempted to prevent *S*'s awareness of errors, several *S*s verbalized awareness of their occurrence. In order to evaluate the possibility

TABLE 2

*F*S AND ERROR *MS*S FROM ANALYSES OF VARIANCE OF GRIP PRESSURE SCORES

Source	*df*	Groups Compared		
		I vs. II	I vs. III	II vs. III
Groups (G)	1	8.31**	6.46*	.57
Stimulus (S)	1	1.98	1.54	.23
G × S	1	2.99	.36	4.27*
Error (b) (*MS*)	52	*11901.04*	*6328.88*	*13388.00*
Trials (T)	5	27.56***	10.99***	40.38***
T × G	5	2.16	.76	.17
T × S	5	.82	.98	1.86
T × G × S	5	.31	.42	.39
Error (w) (*MS*)	260	*442.62*	*1236.28*	*419.19*

* $P = .05$.
** $P = .01$.
*** $P = .001$.

that the present results may have been affected by *Ss'* awareness of errors, grip pressure scores of all *Ss* who made one or more errors were removed and mean pressures based on all trials were recomputed. The resulting means were 23.78 and 22.03 for Groups II and III, respectively. Each of these means was found to be significantly higher than that of Group I at the .01 level when evaluated by *t* tests.

Discussion

The present findings, which are consistent with those reported by Finger with rats, clearly indicate stimulus similarity to be differentially related to the two measures of response strength, latency and grip pressure. More specifically, the latency results are consistent with the interpretation that increasing the similarity between two cues augments generalization between them and thereby heightens competition between the two responses. The increase in grip pressure, on the other hand, provides evidence for the increase in the level of motivation presumed to accompany heightened response competition.

Nonreinforcement and punishment would be expected not only to vary with stimulus similarity but perhaps to contribute to motivation as well; however the measures reported by Finger were obtained prior to *S*'s having experienced such factors. Similarly, the present findings do not support the notion that higher grip pressures may have reflected motivational effects of errors in discrimination. On the other hand, in view of the fact that human *Ss* were employed, the factor of prior experience should be considered. That is, anticipations of poor discrimination performance may have been more highly associated, through prior experience, with the condition involving the less discriminable stimuli. On this basis, the higher grip pressure scores may have reflected heightened tension, concern, apprehension, etc., as a result of anticipation of failure, even though errors did not occur.

As indicated earlier, most related studies have reported weakening, rather than strengthening, of the overt response which has been aroused in competition with an incompatible tendency. It is possible that the conditions in these studies, as contrasted with Finger's and the present one, were such as to render it difficult for the vigor of the overt response to reflect the energizing effect of competition uncomplicated by the inhibiting influence of the competing tendency. That is, the inhibiting influence would seem most likely to persist under conditions which facilitate the continued effectiveness of the cue giving rise to the competing tendency. Thus, it is possible that at the time the selective (jump) response occurred in Finger's situation, the influence of the competing tendency was minimized. Similarly, since the present measure of vigor involved only the terminal segments of the selective, motor response, it may be that the effectiveness of the cue associated with the competing tendency may have been negligible at this point in the response sequence. These suggestions indicate one of the difficulties that may be involved where the response which is initially aroused in competition with another is also the response employed to reflect the increase in motivational level suggested to accompany competition.

Summary

By means of simple motor responses whose latency and vigor (grip pressure) were measured, 84 human *Ss* were required to make brightness discriminations over a total of 48 trials. Different groups discriminated between brightness values separated by differences of either 2.0 or .50 log units. The results showed that latency was reliably longer the smaller the difference between the brightnesses while grip pressures showed the opposite effect, i.e., pressures were reliably higher the smaller the difference. Furthermore, the effect on grip pressure was shown to be independent of errors in discrimination. Among the several interpretations of the findings which were considered was the possibility that conditions which lead to the simultaneous arousal of incompatible responses may serve as a source of drive.

References

BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.

CASTANEDA, A. Reaction time and response amplitude as a function of anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1956, **53**, 225–228.

FINGER, F. W. Quantitative studies of "conflict": I. Variations in latency and strength of the rat's response in a discrimination-jumping situation. *J. comp. Psychol.*, 1941, **31**, 97–127.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Ch. 14.

(Received February 22, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 315–318

PAIRED-ASSOCIATE LEARNING WHEN THE SAME ITEMS OCCUR AS STIMULI AND RESPONSES[1]

ROBERT K. YOUNG

University of Texas

The present paper deals with the the further investigation of a special case of paired-associate (PA) learning. This PA task is one which has been derived from the items of a serial list. That is, a serial list A-B-C-D-E etc., may be thought of as being composed of pairs of items: A-B, B-C, C-D, etc. Random presentation of these pairs B-C, E-F, A-B, etc., wherein all but the first and last items of the serial list serve both as stimulus and response, defines this derived PA task.

Two studies (Primoff, 1938; Young, 1959) have shown that the type of PA learning mentioned above is much more difficult than is a serial list composed of the same items and associations. In addition, it was found (Young, 1959) that when a serial list and a PA list derived from the serial list were learned, high positive transfer occurred from the PA to the serial task but did not occur when the order of learning was reversed. That is, little, if any, positive transfer occurred when transfer was measured from the serial to the PA task.

It would seem that the derived PA list could be viewed as an extreme case of intralist stimulus-response similarity. Since all but two of the items within the list serve both as stimulus and response, the task may represent an identity condition along a dimension of stimulus-response similarity. The typical PA task in which there is low intralist similarity would be at one end of this continuum while the derived PA list in which there are identical stimuli and responses would be at the other end.

At present, no comparison has been made between the learning of a derived PA list as described above and the learning of a PA list constructed in the usual fashion. Although a rough comparison would indicate that a PA list derived from a serial list would be much more difficult than the typical PA list, there is no clear-cut evidence to support such a conclusion.

Intralist stimulus-response similarity has been shown to be directly related to difficulty in PA learning (Underwood, 1951) and if, as suggested above, the derived PA task differs from a typical PA task along a stimulus-response similarity dimension, then the derived PA task should be more difficult than the usual PA task. Such a test is made in the present study. Furthermore, a PA task composed of stimulus-response items which are similar should be more difficult than the usual PA task but less difficult than the derived PA task. This test is also made in the present experiment. Because the influence of similarity may vary as a function of the difficulty of the material as determined by meaningfulness, two levels of meaningfulness have been incorporated in the present design.

METHOD

Design.—The design of the experiment may be thought of as a 2 × 3 factorial design in which the rows represent two degrees of

[1] This work was done under Grant G-4850 from the National Science Foundation.

TABLE 1

LISTS USED IN THE EXPERIMENT

90% lists					
Low		Medium		High	
FAM	LEK	WIP	FUZ	YOK	QIL
COV	RUD	GOL	HIN	GEN	TOS
JAN	BIZ	BIC	DEK	WAV	RUF
SAF	QES	FUT	VOL	QIL	GEN
DEB	WOL	MIN	BAC	TOS	WAV
HIR	PEY	DEY	RAS	RUF	PIC
20% lists					
Low		Medium		High	
YOL	TEF	ZAS	KUY	CIW	ZEV
WEQ	XID	XOH	NAQ	SAJ	LEB
CUH	QAM	GED	TUW	GUH	XOK
KIF	ZUS	LUY	VOH	ZEV	SAJ
NAX	ROJ	NIQ	GEB	LEB	GUH
VIG	BER	TIW	JUF	XOK	RUY

meaningfulness and the columns represent the transition from low to high stimulus-response similarity. The column variable will be designated degree of similarity and will be defined in terms of the formal similarity of the stimulus item of one pair and the response item of some other pair within the list. The Low Similarity condition would simply be a PA list, the Medium Similarity condition would be one in which each stimulus is similar to a response item in another pair (with the exception of one stimulus and one response which are not similar), and the High Similarity condition would be a PA list in which each stimulus is used as the response for another pair within the list (again with the exception of a single stimulus and a single response). Meaningfulness is defined in terms of the association value of nonsense syllables taken from Glaze's lists (Hilgard, 1951).

Subjects.—Seventy-two *Ss* selected from introductory psychology classes learned one six-pair list a day for three successive days. One-half the *Ss* were randomly assigned to the high-meaningfulness (90%) group and the other half were assigned to the low-meaningfulness (20%) group. Each *S* served in all three conditions of similarity with no practice day being given.

Lists and apparatus.—Lists of nonsense syllables were presented by an Animatic filmstrip projector at a standard 2:2-sec. rate with a 4-sec. intertrial interval. The learning criterion for all lists was one perfect recitation. The association values of the lists in the 90% group varied from 91.6 to 93.3% while the mean association value of each list in the 20% group was equal to 20.0%. The three lists learned by a group were counterbalanced, requiring six *Ss* for complete counterbalancing.

The 90% and 20% lists are presented in Table 1. From the table it can be seen that the Low Similarity lists are simply PA lists with low intralist similarity and that the High Similarity lists are PA lists with the same items serving both as stimuli and as responses. Similar pairs of items in the Medium Similarity lists differ by only one letter. For example, the response in the pair GOL-HIN is similar to the stimulus in the pair MIN-BAC. In the Medium Similarity condition both the 90% and 20% lists included pairs which differed either by the first, second, or third letter.

With the exception of the similarity arising from manipulation of stimulus-response similarity, formal intralist similarity was as low as possible in all lists. In the 90% group 6 consonants were used twice in the Low Similarity condition (but never in the same pair) and 3 were used twice in the Low condition of the 20% group. No consonant was used more than once in the Medium and High Similarity conditions of either group with the exception of those which appeared in similar (or identical) stimuli and responses. With reference to the similar and identical items, 7 consonants were repeated in the Medium Similarity conditions and 10 in the High Similarity conditions. Thus the formal similarity between conditions was approximately equivalent. No vowel was used more than four times in any of the lists in either group.

Three orders of presentation were used for each of the lists in both groups to reduce the possibility of serial learning. In addition, pairs which had items (or similar items in the case of the Medium Similarity conditions) in common were never presented in succession.

RESULTS AND DISCUSSION

Trials to learn.—The mean numbers of trials to learn as a function of meaningfulness and degree of similarity are presented in Table 2. The variances obtained were found to be heterogeneous and the data have not been transformed.

Nonparametric techniques were em-

ployed to test for the significance of differences. The influence of meaningfulness was analyzed by use of the Mann-Whitney *U* test (Siegel, 1956, pp. 116–127). With *N*s of 36 and 36 the obtained *U* was transformed into a *z* score of 6.08 which is significant beyond the .001 level, indicating that the 90% lists were easier to learn than the 20% lists. Similarity conditions were analyzed separately for each group by use of the Friedman two-way analysis of variance (Siegel, 1956, pp. 166–172). The χ_r^2 obtained for the 90% group was found to equal 38.38 and for the 20% group $\chi_r^2 = 26.72$. With 2 *df* a χ^2 of 13.82 is required for significance at the .001 level indicating that difficulty increased as similarity increased. Thus, both meaningfulness and similarity had a significant influence upon number of trials to criterion.

Additional comparisons were made between the High and Medium Similarity conditions in each group. The χ_r^2 obtained between these two conditions in the 90% group was found to equal 12.25 which is significant beyond the .001 level with 1 *df*, but the χ_r^2 obtained between the same two conditions of the 20% group was found to equal 1.00 which does not reach significance. The difficulty of the High Similarity list in the 20% group probably tends to be underestimated since 12 of the 18 *S*s dropped from the experiment were lost because of failure to learn this list. This loss of *S*s may account for the absence of a significant difference between the High and Medium conditions of the 20% group.

TABLE 2

TRIALS TO LEARN TO CRITERION

Meaningfulness	Similarity					
	Low		Medium		High	
	Mean	σ_m	Mean	σ_m	Mean	σ_m
90%	11.1	1.04	15.8	1.68	22.6	1.69
20%	24.2	1.83	36.1	2.05	37.4	2.13

TABLE 3

OVERT ERRORS PER TRIAL TO CRITERION

Meaningfulness	Similarity					
	Low		Medium		High	
	Mean	σ_m	Mean	σ_m	Mean	σ_m
90%	.72	.06	.83	.08	1.27	.09
20%	.77	.06	.73	.06	.79	.06

The present findings indicating that learning of paired nonsense syllables is a function of meaningfulness is consistent with other research, e.g., Noble and McNeely (1957), in that the higher the meaningfulness of the verbal material the faster the learning.

The data also indicate that difficulty increases as similarity increases. Since the mean of the High Similarity PA list in the 20% group tends to be underestimated because of the loss of *S*s, it may be that the lower the meaningfulness the greater the influence similarity has in retarding rate of learning. Although this would be consistent with serial-learning data in which meaningfulness and intralist similarity were manipulated (Underwood & Richardson, 1956) no unequivocal statement can be made in the present experiment concerning the interaction of meaningfulness and similarity. In any event it is clear that similarity as manipulated in the present experiment is a potent variable and that the difficulty of derived PA learning is largely attributable to the similarity conditions within the list.

Errors.—The overt error data were analyzed in terms of overt errors per trial and are presented in Table 3. Overt errors per trial were found to be a function of the conditions of Similarity ($F = 19.10$, $df = 2/136$, $P < .01$), Meaningfulness ($F = 5.05$, $df = 1/70$, $P < .05$), and the interaction between Meaningfulness and Similarity ($F = 14.65$, $P < .01$). In addition, separate analyses of the 90% and 20% groups indicate that errors per trial were a function of similarity when meaningfulness was high ($F = 29.08$, $df = 2/68$, $P < .01$) but not when meaningfulness was low ($F < 1.00$). Thus while the 90% lists were learned faster, more errors per trial were elicited during the learning.

When meaningfulness was high, list difficulty and error frequency were directly related to intralist similarity. When meaningfulness was low, difficulty increased as intralist similarity increased but error frequency did not. Like many previous studies there is a lack of a consistent relationship between error frequency and learning as intralist similarity varies. Underwood (1953a, 1953b), for example, found intralist response similarity related to error frequency but not to rate of learning and he also found intralist stimulus similarity tended to be related to learning but not to error frequency. No immediately apparent explanation is available for the lack of a systematic relation between similarity, meaningfulness, rate of learning, and error frequency obtained in the results of the present experiment.

Summary

In the present experiment the formal similarity between the stimulus of one pair of nonsense syllables and the response of another pair was varied. Under conditions of low stimulus-response similarity a paired-associate (PA) task with low intralist similarity was learned. Under conditions of high stimulus-response similarity a PA task was learned in which all but two items in the six-pair list served both as the response for one pair and the stimulus for another. In terms of formal similarity all lists were approximately the same. Two groups of lists were learned, one group having high meaningfulness (90% association-value nonsense syllables) and the other having low meaningfulness (20% association-value nonsense syllables).

The major findings of the experiment were that rate of learning was directly related to meaningfulness and inversely related to degree of stimulus-response similarity. In addition it was found that frequency of overt errors was directly related to similarity under conditions of high meaningfulness but was not related under conditions of low meaningfulness.

References

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 517–567.

Noble, C. E., & McNeely, D. A. The role of meaningfulness (*m*) in paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

Primoff, E. Backward and forward associations as an organized act in serial and in paired associate learning. *J. Psychol.*, 1938, **5**, 375–396.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.

Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–143. (a)

Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149. (b)

Underwood, B. J., & Richardson, J. The influence of meaningfulness, intralist similarity, and serial position on retention. *J. exp. Psychol.*, 1956, **52**, 119–126.

Young, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

(Received March 7, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 319-321

THE EFFECT OF CONTINUOUS, PARTIAL, AND VARIED MAGNITUDE REINFORCEMENT ON ACQUISITION AND EXTINCTION

HARRY G. YAMAGUCHI

Indiana University

Logan, Beier, and Ellis (1955) initiated the study of varied magnitude reinforcement wherein all training trials are reinforced but with variations in amount from trial to trial. Their predictions of the effect of this type of reinforcement schedule on acquisition and extinction (Logan, Beier, & Kincaid, 1956) were derived from a micromolar extension (Logan, 1956) of Hullian theory (Hull, 1952; Spence, 1956). The theory of Logan et al., when applied to various reinforcement schedules identical in total amount of reinforcement but different in variability of amount, would predict that (*a*) in acquisition performance, the effect of varied magnitude reinforcement will be similar to that of continuous reinforcement but inferior to the effect of partial reinforcement, and (*b*) in resistance to extinction, the effect of varied magnitude reinforcement will be intermediate to those of partial and continuous reinforcement.

The present experiment is a further study of the relative effects of varied, partial, and continuous reinforcement on acquisition and extinction.

METHOD

Subjects and apparatus.—The *S*s were 52 male hooded rats bred in the Indiana University animal colony and were 90 to 120 days old at the beginning of experimentation. The apparatus was a runway 48 in. long and 4½ in. wide with a starting box 8 in. long and 6 in. wide and a goalbox 12 in. long and 8 in. wide. The walls (9 in. high) and floor were made of ¼-in. plywood painted flat black and the roof consisted of hinged Plexiglas lids. There were two guillotine doors in the runway, one between the starting box and the runway and the other 6 in. from the end of the runway. A food cup 1 in. square and ¼ in. deep was attached to the back wall of the goalbox, 3 in. above the floor. The time interval required to traverse 46 in. of the runway was measured by a Standard Electric timer which was started by raising the starting box door and was stopped by a photocell located 2 in. in front of the goalbox. The apparatus was illuminated by a 60-w. ceiling light 4 ft. above the center of the apparatus.

Procedure.—The *S*s were housed singly and were maintained on daily rations of 15 gm. of mash (42% ground Purina laboratory chow, 14% flour, and 44% water by weight) with water always available. On acquisition days, *S*s were fed 2½ gm. of mash in the apparatus and 12½ gm. in their home cages.

During 10 days of habituation, *S*s were handled, in groups of 4, for 10 min. per day. In addition, on Days 8 and 9 of habituation, *S* was allowed to explore the starting box and runway and on Day 10, *S* was fed its daily ration in the goalbox with the goalbox door closed. At the end of habituation, *S*s in each habituation group were randomly assigned to a reinforcement schedule, one *S* per schedule.

The four reinforcement schedules were: (1) 5–5 continuous reinforcement, 5 units of reinforcement on all trials; (2) 8–2 varied magnitude reinforcement, 8 units on a random half of the trials and 2 units on the other half; (3) 9–1 varied magnitude reinforcement, 9 units of reinforcement on a random half of the trials and 1 unit on the other half; and, (4) 10–0 partial reinforcement, 10 units on a random half of the trials and no reinforcement on the other half. A unit of reinforcement was .05 gm. of mash. For each *S* in the partial and varied reinforcement groups, the sequence of 5 larger and 5 smaller reinforcements per 10 trials was determined randomly with the restriction of reinforcing the first trial of each day with the larger amount and of setting a limit of three successive reinforcements of the same amount.

The *S*s were given 40 acquisition trials, 10 trials per day with an intertrial interval

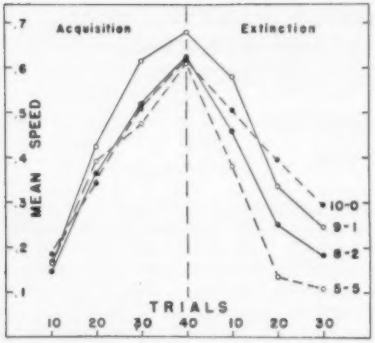

FIG. 1. Mean running speed in 10-trial blocks for the various groups.

of 25 sec. Reinforcements were in the form of a single pellet of mash regardless of the number of units involved. On all trials, *Ss* were removed from the goalbox 15 sec. after interrupting the photobeam.

On the day following the last acquisition trial, *Ss* were given 30 massed extinction trials with a 25-sec. intertrial interval and a 15-sec. goalbox confinement period. If *S* failed to enter the goalbox within 120 sec., it was removed from the apparatus and if such failures occurred on 5 consecutive trials, extinction was discontinued. For no-response and discontinued trials, runway latency was regarded as indeterminate and reciprocal latency as zero.

RESULTS

All runway latencies were converted into speed (reciprocal latency) measures. The mean speed of the four groups on 10-trial blocks of acquisition and extinction is shown in Fig. 1.

Table 1 presents a summary of the acquisition data for two overlapping portions of acquisition training. The total acquisition data are based on each *S*'s mean speed on 40 trials and the terminal acquisition data on each *S*'s mean speed on the last 5 trials. An analysis of variance of the total acquisition data revealed no significant differences ($F < 1$). Similar results were obtained in an evaluation of the terminal acquisition data which, because of nonhomogeneity of variance, were tested by a Kruskal-Wallis one-way analysis of variance ($H = 2.53$, $P > .30$).

TABLE 1

ACQUISITION SPEED DATA

Group	Total Acquisition		Terminal Acquisition	
	Mean	*SD*	Mean	*SD*
5–5	.411	.200	.611	.208
8–2	.413	.118	.637	.116
9–1	.471	.085	.708	.091
10–0	.414	.167	.629	.218

Two measures of extinction speed are summarized in Table 2. The mean extinction speed measure is based on each *S*'s mean speed on 30 extinction trials. The median relative speed measure is similar to the extinction measure used by Logan et al. (1956) and was determined, for each *S*, by computing the ratio of (*a*) median extinction speed minus terminal extinction speed (mean of last 5 extinction trials) to (*b*) terminal acquisition speed (mean of last 5 acquisition trials) minus terminal extinction speed. An analysis of variance of the mean extinction speed data yielded an *F* significant at the 2.5% level. Application of the Tukey test (Snedecor, 1956) indicated a significant difference ($P < .05$) between the partial and the continuous reinforcement group and between the 9–1 varied magnitude and the continuous group. An analysis of the

TABLE 2

EXTINCTION SPEED DATA

Group	Mean Extinction		Median Relative Extinction	
	Mean	*SD*	Mean	*SD*
5–5	.208	.126	.027	.110
8–2	.298	.148	.273	.445
9–1	.388	.143	.318	.432
10–0	.398	.212	.526	.472

median relative speed data produced identical results; a Kruskal-Wallis test of the four groups yielded a significant H ($P < .05$) and the White T test revealed significant differences between the 10-0 and 5–5 groups ($P < .01$) and between the 9–1 and 5–5 groups ($P < .05$).

Discussion

The extinction results of the present study are consistent with the results obtained by Logan et al. (1956) and are in accord with their theory. The acquisition data, however, do not support their theory.

For acquisition performance, their theory (Logan et al., 1955) holds that different amounts of reinforcement produce different fractional anticipatory responses (r_g's). In varied magnitude reinforcement, the different r_g's compete and result in a compromise r_g equivalent to the r_g of the mean amount of reinforcement. On the other hand, in partial reinforcement, no small r_g is produced to compete with the large one. Their theory, therefore, predicts that partial reinforcement should produce better acquisition performance than varied magnitude reinforcement. Although the results of a supplementary group in their study confirmed this prediction, the results of a second experiment (Logan et al., 1956) did not. In the latter experiment, the asymptotic speed of a 9-1 varied magnitude group was somewhat higher than that of a 9–0 partial reinforcement group. With this result and the results of the present study, it seems reasonable to conclude that partial, varied magnitude, and continuous reinforcement schedules produce similar acquisition performance when total amount of reinforcement is held constant. This conclusion, however, suggests that some modification of the theory of Logan et al. is necessary.

The adequacy of a simple alteration, such as the assumption that partial reinforcement also produces a compromise r_g, remains an experimental question. The finding of Leventhal, Morrell, Morgan, and Perkins (1959) that rats exhibit a preference for partial reinforcement over continuous mean magnitude reinforcement if the reward is small but not if it is large suggests that the amount of reinforcement must be considered as well as the variability of reinforcement.

Summary

The effects of continuous, partial, and varied magnitude reinforcement on acquisition and extinction of a runway response was studied with four groups of 13 rats. Each group was given one of the following reinforcement schedules: (*a*) continuous reinforcement, 5 units on all trials; (*b*) 8–2 varied magnitude, 8 units on a random half and two units on the remainder of the trials; (*c*) 9–1 varied magnitude, 9 units on a random half and one unit on the other half of the trials; and, (*d*) partial reinforcement, 10 units on a random half of the trials and no reinforcement on the other half. All *Ss* were given 40 acquisition and 30 extinction trials.

The various groups exhibited similar acquisition performance but showed significantly different extinction effects. The partial reinforcement and the 9–1 varied magnitude groups were more resistant to extinction than the continuous reinforcement group. The theoretical implications of these results on a micromolar theory of the fractional anticipatory goal response were discussed.

References

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Leventhal, A. M., Morrell, R. F., Morgan, E. F., Jr., & Perkins, C. C., Jr. The relation between mean reward and mean reinforcement. *J. exp. Psychol.*, 1959, **57**, 284–287.

Logan, F. A. A micromolar approach to behavior theory. *Psychol. Rev.*, 1956, **63**, 63–73.

Logan, F. A., Beier, E. M., & Ellis, R. A. Effect of varied reinforcement on speed of locomotion. *J. exp. Psychol.*, 1955, **49**, 260–266.

Logan, F. A., Beier, E. M., & Kincaid, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

Snedecor, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State Coll. Press, 1956.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received March 9, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 322–328

ACQUISITION AND TRANSFER IN PATTERN-VS.-COMPONENT DISCRIMINATION LEARNING

W. K. ESTES[1] AND B. L. HOPKINS

Indiana University

Common sense, fortified by a few experimental observations (Hull, 1943; Pavlov, 1927), would lead us to expect that learning to discriminate a stimulus pattern from one or more of its components should proceed in much the same manner as learning to discriminate two entirely dissimilar patterns. The discrimination of *ab* from *a*, say, should be no more difficult than the discrimination of *ab* from *cd*. Contemporary statistical models for discrimination learning, in their simplest form (Burke & Estes, 1957), predict however, that learning should be more difficult in the former instance, and in fact that the limit of accuracy in discriminative performance should be inversely related to the proportion of cues (components) common to the two patterns undergoing discrimination. In the case of *ab* vs. *a*, e.g., the expected asymptotic probability of correct responding would be less than unity (only .625 if the two patterns occurred equally often during the series of differentially reinforced training trials).

We have no doubt that at least for human *S*s, and quite possibly for all animals that demonstrate discrimination learning at all, common sense is right and the model is wrong on this point. Still, an analysis of the problem in terms of the model raises a number of questions to which neither common sense nor the experimental literature provides satisfactory answers. Within a pattern-vs.-component discrimination, how will learning proceed on the pattern and the components considered separately? According to some current interpretations the attainment of 100% accuracy on discrimination problems involving overlapping stimuli depends upon the acquisition of an "observing response" (Atkinson, 1958; Wyckoff, 1952) or the equivalent (Restle, 1955); that is, *S* learns to ignore common cues and attend only to those that are unique to one or the other of the discriminative stimuli. In the discrimination of *ab* vs. *a*, *S* learns to respond to the *ab* pattern as a unit, in some sense, but during the course of learning, he should exhibit negative transfer effects from the pattern to the component and vice versa. An alternative view is that the pattern is responded to as a unit from the outset; if so, then an asymptote of 100% accuracy in discrimination between *ab* and *a* can again be predicted, but there should be no negative transfer effects during learning. One purpose of the present study is to obtain quantitative data on learning rates, asymptotes, and transfer effects which may be useful in evaluating these alternative interpretations. A second purpose has to do with some of the specific assumptions of statistical learning theory.

Anticipating that the statistical model developed in relation to simpler learning situations (Burke & Estes, 1957) will prove inadequate, we are interested in seeing whether it can be revised or extended to handle pattern-vs.-component discrimination. The shortcomings of the model may lie either in the assumptions or in the customary interpretations. The principal assumptions are (*a*) that the probability of a given response to any element of a stimulating situation increases by a linear transformation of its

[1] This paper was prepared while the first author held a visiting appointment in the Department of Psychology, Northwestern University, Spring Quarter, 1959.

previous value upon each reinforcement, and (*b*) that probability of a response to a stimulus compound is a weighted average of its probabilities to the constituent elements separately. These assumptions have accumulated considerable empirical support in a variety of contexts (Estes, 1959). Consequently a natural strategy is to see whether a model embodying these assumptions may account for the problem at hand if we reinterpret appropriately the rules of correspondence between terms of the model and terms of the empirical situation.

In particular, it has been the practice in previous applications of the statistical model (e.g., Estes, Burke, Atkinson, & Frankmann, 1957; Schoeffler, 1954) to treat the stimulus elements of the model as the conceptual counterparts of the separably manipulable components of the stimulating situation. In this interpretation, Components *a* and *b* in the *ab* vs. *a* discrimination would be considered to represent stimulus elements and the response probability to *ab* would be the average of the probabilities to *a* and *b* separately. Possibly these identifications are wrong, and it is not the separate components that act as elements for the learner. What, then, does act as an element? Several possible answers are suggested by considerations arising from contemporary learning theory. In a later section these will be summarized and checked out against our data. In order to be able to differentiate among these possibilities, we have included in our experimental design a number of tests on which components of previously discriminated patterns appear in new combinations. By giving such tests at both intermediate and terminal stages of learning, we may be able to determine precisely what aspects of the stimulating situation, if any, are acting as stimulus elements in the sense required by statistical learning theory.

TABLE 1

PARADIGM FOR DISCRIMINATION TRAINING

Relative Frequency	Stimulus	Reinforced Response
1/8	*a*	1
1/8	*b*	2
1/4	*ab*	1
1/8	*c*	1
1/8	*d*	2
1/4	*cd*	2

METHOD

Subjects and apparatus.[2]—The *S*s were 45 students from undergraduate psychology classes at Indiana University.

Four booths, separated by plywood panels, were located in the experimental room. On a panel at the front of each booth were mounted two 6-v. pilot lights, one above the other and 6 in. apart, and a lever which could be moved up or down from its normal center position.

A black screen was positioned about 8 ft. in front of the booths, well within *S*s' lines of vision when they were seated in the booths. A slide projector was used to present stimulus figures on the screen. The durations and intervals between presentations of the stimuli and the pilot lights were controlled by a programer made from a modified Esterline-Angus recorder. Another Esterline-Angus recorder was used to record *S*s' lever responses.

The figures shown on the screen were various combinations of four Russian letters, which will henceforth be designated *a*, *b*, *c*, and *d*, respectively. The letters projected on the screen as figures about 3½ in. high and appeared black against a lighter background.

Experimental design.—Each *S* was given training concurrently on two pattern-vs.-component discriminations according to the paradigm exhibited in Table 1. The right hand column of the table gives the response reinforced in the presence of each of the stimulus patterns shown in the middle column. By "pattern" is denoted simply the complete stimulus figure shown on a trial, whether it consists of one or more than one

[2] We are indebted to David LaBerge for making the apparatus available to us and to Marcia Johns for the use of the slides.

letter. In the case of the double letters, *ab* and *cd*, the spatial positions of the two constituent letters were the same throughout training. Reinforcement of the upward or downward lever movement consisted in illumination of the upper or lower pilot light, respectively. The two directions of lever movement were assigned randomly to the roles of "Response 1" and "Response 2" in the paradigm. The reason for giving two concurrent problems was to prevent the number of letters per se from being differentially correlated with reinforcement of either response.

Test trials, given midway through training and again at the end, consisted in presentation of 37 different letter combinations. These were: all 12 ordered pairs, 19 triads, and 6 permutations of all four letters. Two random orders of the test patterns were used, one for 24 of the *S*s and the second for the remainder. Since there proved to be no noteworthy differences in results between these subgroups, their data will be pooled.

Procedure.—The *S*s were seated one to a booth, in full sets of four so far as possible, and *E* read the following instructions:

> I want to see how well you can do in a rather unusual problem situation. At the beginning of each trial you will see a figure on the screen for about two seconds. About 2 sec. after you see a figure, either the top or the bottom light in your booth will go on. While the figure is being shown on each trial, you are to show whether you expect the top light or the bottom light to come on by pushing the black lever up or down. . . . If you are not sure, guess. Be sure to make your choice while the figure is visible. It is important that you push the lever up or down, never both, on each trial.
>
> During the first few dozen trials, either the upper or the lower light in your booth will flash to show you which way was the correct response to the pattern shown on the screen. Then we will begin having some test trials on which neither of your booth lights will flash. These test trials are included to see if you can keep up the same level of accuracy when you are not being told whether you are right or wrong on every trial. Be sure to continue making your best guess on every trial.

On each training trial, a pattern was projected on the screen for 2 sec.; there was an empty interval of 2 sec.; the pilot light corresponding to the pattern shown was lighted for 2 sec.; then there was another empty interval of 2 sec. before the next pattern was shown. The order of the training patterns was randomized within successive cycles of 8 trials (a single order for all *S*s), each cycle including each single letter once and each of the pairs *ab* and *cd* twice. Test trials were conducted just as the training trials except that the pilot-light interval was omitted. A session began with 8 cycles of training trials, then a block of 37 test trials, 8 more cycles of training trials, and a final block of 37 test trials.

Results and Discussion

Rate and final level of discrimination learning.—As anticipated, *S*s were approaching 100% accuracy in discriminative performance by the end of the second training series. The proportion of correct responses over the last cycle of eight training trials was .94, and since it was still rising, this proportion would undoubtedly have approximated 1.0 with sufficient additional training. Considering responses to the single letters and the two-letter patterns separately, the proportions correct were .81 and .87, respectively, at the end of the first training series and .92 and .96 at the end of the second. Probably the best single comparison, equating for number of reinforcements and avoiding any complications due to forgetting over the first test series, is that between the proportions of correct responses over the first eight training trials on each stimulus; these values are .68 and .71 for the single and double stimuli, respectively. Thus we find no shadow of evidence that learning was slower on the two-stimulus patterns than on the singles.

Transfer effects during discrimination learning.—If transfer occurs from patterns to components, then we should expect more errors on Components *b* and *c* than on *a* and *d*, since the response correct for *ab* is incorrect for *b* and the response correct for *cd* is incorrect for *c*.

Computation of the mean number of errors per *S* during learning on each component yields 4.56 and 3.82 for *b* and *c*, 2.33 and 2.84 for *a* and *d*, respectively. Mean trials to the first correct response on each component in each series exhibit exactly the same pattern. Apparently we may conclude that a response learned to a pattern tends to transfer to constituents of the pattern when they appear alone.

Transfer to new test patterns.—In Table 2 are assembled the two and three-letter patterns tested, the proportion of "correct" responses to each, and the number of observations on which each proportion is based. In the case of Test Patterns *ac* and *bd*, the correct responses are of course 1 and 2, respectively. For the patterns of three letters including both *a* and *b*, Response 1 is considered correct, and for those including both *c* and *d*, Response 2 is considered correct. The patterns of four letters yielded Response 1 proportions ranging from .42 to .49, and since differences among them bear no apparent relation to the ordering of the components, they are not reported individually.

The Training Patterns *ab* and *cd* were included in each test series to provide a means of estimating the amount of retention loss during the test series. In both test series these retention tests yielded decrements in correct response proportions from the last training cycle. To quantify the decrement, we may assume that response probability on a test trial is a weighted average of the probability at the end of training and the chance value that would obtain if forgetting were complete. Over the last cycle of the first training series, the proportion correct was .84, and the proportion on the test was .74. Letting r be our coefficient of retention, and assuming that test probability would be .5 if forgetting were complete, we can write $.74 = .84r + .5(1 - r)$, yielding $r = .7$ for the desired estimate. Similarly, for the second test series, we obtain $r = .6$.

TABLE 2

PROPORTIONS OF CORRECT RESPONSES TO TEST PATTERNS

Pattern	*N*	First Test *P*	Second Test *P*
ab, cd	90	.74	.76
ba, dc	90	.66	.58
ac, bd	90	.72	.79
ca, db	90	.67	.76
abc, cab, bcd, cdb	180	.64	.73
bca, bac, bdc, dcb, dbc	225	.60	.65
abd, dab, acd	135	.38	.39
bda, bad, adb, dba, dca, dac, cad	315	.47	.39

The remaining types of test combinations shown in Table 2 were designed to differentiate between two interpretations which may for brevity be designated the *component hypothesis* and the *pattern hypothesis*, respectively. According to the former, the probability of a response to any given component should tend to approach the marginal relative frequency with which the response is reinforced over all training trials on which the component occurs, whether alone or in combination with others. From Table 1 one can readily calculate that for Response 1 and Components *a* and *b*, these relative frequencies are 1.0 and .67, respectively, and similarly for Response 2 in relation to Components *d* and *c*, respectively. These are, then, the values which should be averaged in order to predict the asymptotic probabilities of correct responses on test trials. For Test Pattern *ba* the predicted value is $.5(1.0 + .67) = .83$, for *ac* (recalling that Response 1 is des-

TABLE 3

THEORETICAL ASYMPTOTIC PROBABILITIES OF CORRECT RESPONSES TO TEST PATTERNS, CORRECTED FOR RETENTION LOSS

Pattern	Component Hypothesis	"Molecular" Pattern Hypothesis	"Molar" Pattern Hypothesis
ba	.71	.50	.50
ac	.61	.82	.82
abc	.61	.61	.82
bca	.61	.61	.61
abd	.53	.39	.50
bda	.53	.39	.39

ignated "correct") $.5(1.0+.33)=.67$, for *abc* $.33(1.0+.67+.33)=.67$, for *abd* $.33(1.0+.67+0)=.55$, and so on.

According to the pattern hypothesis, the probability of a response to any component or combination of components should tend to approach the relative frequency with which the response is reinforced during training on trials when that component or combination occurs as a pattern. From Table 1 it is apparent that these relative frequencies for Response 1 and Patterns *a*, *b*, and *ab* are 1, 0, and 1, respectively, and it is these values that should be averaged to predict correct response probabilities to test patterns. For Test Pattern *ba*, the predicted value is .5, for *ac* 1.0; for *abc* it is .67 or 1.0 depending on whether *a* and *b* separately or *ab* as a unitary pattern take the role of elements in mediating the transfer, and similarly for *abd* it is either .33 or .5.

If one of these hypotheses is substantially correct, then the effect of the two training series should be to move the test probabilities progressively away from the original chance level and toward the predicted asymptotic values. It should be noted that in the present experiment the test probabilities cannot be expected to reach the appropriate asymptotic values, for discrimination training was terminated somewhat short of asymptote and also there was some retention loss over the test series. Having available the results of the retention tests with the training patterns, we can, however, correct the theoretical test probabilities obtained above for retention loss. The "coefficient of retention" r was found to be .7 and .6 for the first and second test series; the mean of these two values has been used in computing the corrected theoretical values shown in Table 3 (corrected value = uncorrected value $\times .65 + .5 \times .35$). For brevity, the first entry under "Pattern" in any row of Table 2 is used to designate the corresponding row of Table 3. Only six rows are needed in Table 3 since all three hypotheses yield the same predictions for the third and fourth rows of Table 2. The labels "molecular" and "molar" in the headings of the last two columns of Table 3 refer to the two forms of the pattern hypothesis; according to the former, *a* and *b* (or *c* and *d*) separately should behave as elements in mediating transfer to test triads, whereas according to the latter the combination *ab* (or *cd*) should behave as an element.

Consideration of the first two rows of Table 3 is sufficient to dispose of the component hypothesis. Comparing these with the corresponding *ba* and *ac* rows of Table 2, we see that the test probabilities differ considerably from the predicted values on the first test and move further away on the second test in each instance.

On the other hand, the data for the test pairs appear quite compatible with the pattern hypothesis. In each case the observed proportion moves substantially in the direction of the predicted asymptote between the first and second test and is about as

close on the second test as could be expected, considering that the final level of correct responding on the single letters during training was only .92. On test triads of the type *bca* and *bda*, where members of a training pair do not appear in their original arrangement, predictions are again the same for both versions of the pattern hypothesis. For the former, the observed values of .60 and .65 are quite close to the theoretical .61; for the latter, the first test result of .47 is far short but the second test result of .39 agrees with the common prediction.

To differentiate between the "molecular" and "molar" forms of the pattern hypothesis, we must turn to the remaining two test triads. In the case of type *abc*, both observed proportions fall between the two theoretical values, but the change from the first to the second test is distinctly in the direction of the "molar" outcome. In the case of type *abd*, the first test result agrees well with the "molecular" prediction; the change from the first to the second test is in the direction of the "molar" outcome, but the shift is small and does not approach statistical significance ($SE_p = .04$ on each test). Apparently in these ambiguous cases either *a* and *b* separately or the combination *ab* may be responded to as units, but the likelihood of the latter result increases relative to the former as training progresses. The difference in amounts of change between first and second tests for the triads *abc* and *abd* is quite in line with this interpretation. For *S*s who had not mastered the discrimination at the time of the first test, and who thus still had chance probabilities of correct response to some or all of the training stimuli, continued learning relative either to the separate components or to the two-letter pattern would lead to further increase in the probability of correct response to Test Pattern *abc*. In the case of *abd*, however, learning relative to the separate components would tend to move the test probability downward toward the "molecular" asymptote while learning relative to Pattern *ab* would tend to hold the test probability at .5.

It should be remarked that our data provide no grounds for assuming that the test probabilities specified by the "molar" form of the pattern hypothesis would in all cases be reached with sufficient training. Asymptotic probabilities of response to ambiguous test patterns may well depend on the relative frequencies with which the various subpatterns occur during training.

Theoretical implications.—With respect to the two assumptions of statistical learning theory mentioned earlier, both the "linear operator" representing effects of reinforcement and the "averaging rule" for response to new stimulus compounds receive quantitative confirmation, once it has been determined precisely what aspects of the stimulating situation function as "elements" under the type of experimental arrangement used in this study. These assumptions, appropriately interpreted, apparently account for the discrimination asymptotes, the relative rates of learning on patterns and components, and the probabilities of response to test patterns. One qualification needs emphasis, however.

In previously published applications of the theory to discrimination learning (e.g., Burke & Estes, 1957; Estes et al., 1957), it has been assumed that the probability p_{hi} of response i to stimulus component h changes in accordance with the function

$$p_{hi,n+1} = (1 - \theta)p_{hi,n} + \theta,$$

$(0 < \theta \leq 1)$, upon each occasion when response i is reinforced in the presence

of component *h*, whether *h* occurs alone or in combination with other components. The transfer effects observed during learning indicate that this assumption may hold during the earlier part of discrimination training even in the pattern-vs.-component situation. But once *i* has been conditioned to *h* by virtue of reinforcement on trials involving *h* alone as the stimulus pattern, the probability of response *i* to *h* alone is unaffected by the reinforcement of other responses on later trials when *h* occurs as a component of a larger pattern.

Summary

Training was given human *Ss* concurrently on two problems requiring the discrimination of stimulus patterns from subpatterns. Midway through the learning series and again at the end, nonreinforced test trials were given with new compounds of the training stimuli.

During the learning series, the probability of a given response to any stimulus pattern or subpattern was found to approach the relative frequency with which it had been reinforced in the presence of that stimulus alone. Results of the test trials support the assumption that probability of a response to a novel compound is a weighted mean of its probabilities to the constituents. When the test compound is unambiguous, i.e., none of the training patterns appears in it intact, the response probability is simply the arithmetic mean of the component probabilities. When the test compound is ambiguous, i.e., one of the training patterns appears as a component of the test compound, *Ss* may respond either to the training pattern or to its subpatterns as units, the pattern apparently becoming increasingly dominant over its constituents as training progresses.

References

Atkinson, R. C. A Markov model for discrimination learning. *Psychometrika*, 1958, **23**, 309–22.

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–45.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

Estes, W. K., Burke, C. J., Atkinson, R. C., & Frankmann, J. P. Probabilistic discrimination learning. *J. exp. Psychol.*, 1957, **54**, 233–39.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

Schoeffler, M. S. Probability of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1954, **48**, 323–29.

Wyckoff, L. B., Jr. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–42.

(Received March 7, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 329–333

CONCEPT IDENTIFICATION AS A FUNCTION OF INTRA- AND INTERDIMENSIONAL VARIABILITY[1]

WILLIAM F. BATTIG AND LYLE E. BOURNE, JR.

University of Virginia *University of Utah*

In psychology, the phrase "concept formation" has been used typically to refer to behavior reflecting a categorization or classification process characterized by (*a*) identical responses given to a variety of stimuli which are alike or similar in some features or attributes and (*b*) different responses given to stimuli which are dissimilar in one or more crucial respects. Laboratory investigations of concept formation usually have involved the simultaneous learning of several mutually exclusive concepts by requiring *S* to sort a large array of stimuli into a few categories on the basis of similarities and differences among them. One example of a relatively standardized procedure of this type is the "concept identification" task (Archer, Bourne, & Brown, 1955) which has been used to demonstrate the effects of several important variables and has, in addition, formed the basis for the extension of a mathematical theory of discrimination learning to concept formation (Bourne & Restle, 1959).

One shortcoming of the concept identification task, and of many other procedures used in experiments purportedly concerned with concept formation, is that in a real sense they do not require that concepts be formed. This is because all of the stimuli to be classified with the same response are identical with respect to the crucial or relevant characteristics. For example, if the concept is "red," all stimuli are exactly the same red, or if "triangle," all stimuli are triangles of identical shape. As a result, the task can be learned in the same way as any discrimination learning problem, differing only in the greater difficulty of finding the basis for the discrimination or discriminations to be learned. The classification of stimuli which are not identical in any way, but only similar, as is characteristic of most concepts, is for the most part absent from these tasks. In fact, an effort is typically made to select dimensions which do not interact with each other, in order to further reduce variations of this nature.

The present experiment represents an attempt to incorporate into the concept identification task variations of an *intradimensional* nature, so that, for example, the concept "red" represents different degrees or shades of red instead of just a single red. Task complexity was varied systematically in terms of the extent of intradimensional differences as well as interdimensional variation (number of irrelevant dimensions) which has been thoroughly investigated in previous studies (Bourne, 1957; Bourne & Haygood, 1959). The results permit a comparison of the two methods of varying task complexity and the determination of any interaction be-

[1] This research was carried out at the University of Utah and was supported in part by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service, with the second author as principal investigator. The authors wish to acknowledge the contributions of R. C. Manion and R. C. Haygood, who assisted in the collection and analysis of the data.

TABLE 1

DESCRIPTION OF EXPERIMENTAL CONDITIONS AND DESIGN

Dimension	Relevance to Solution	Stimuli Used for Successively Increasing Intradimensional Variability						% Between-Category Errors
		2 Values		4 Values		6 Values		
		Category A	Category B	Category A	Category B	Category A	Category B	
Form	relevant	right triangle	square	equilateral triangle	trapezoid	obtuse triangle	parallelogram	2.9
Width	relevant	4/7 in.	1 5/16 in.	1/2 in.	1 1/2 in.	7/16 in.	1 5/7 in.	7.2
Color	irrelevant for all cond.	medium red	medium green	light red	light green	dark red	dark green	1.7
Position	irrelevant for 2 & 3 irrel. dim. cond.	1 in. left of center	1 in. right of center	1 1/2 in. left of center	1 1/2 in. right of center	2 in. left of center	2 in. right of center	2.5
Stripes	irrelevant for 3 irrel. dim. cond.	1 horiz.	1 vert.	2 horiz.	2 vert.	3 horiz.	3 vert.	5.0

Note.—In reading this table, it will be helpful to regard both the rows and columns as cumulative, such that each condition also includes those dimensions and conditions described above and to the left of it.

tween them. In addition, they provide the basis for a possible extension of the Bourne-Restle mathematical model.

METHOD

Apparatus.—The apparatus was identical to that used in previous studies of concept identification, and has been described in detail elsewhere (Bourne, 1957). Geometric stimulus patterns were projected on a screen in front of *S* who responded under completely self-paced conditions by pressing one of four buttons on a response panel. A noncorrection procedure was used such that, immediately following *S*'s response, the light corresponding to the correct button on the response panel was lighted for 1 sec., and was followed 4 sec. later by the presentation of a new stimulus pattern. The *S*'s responses were recorded automatically on an Esterline-Angus operations recorder, as was the correct response for each pattern.

Design.—A 3×3 factorial design was employed, representing all possible combinations of three conditions of interdimensional variability (1, 2, and 3 irrelevant dimensions) and three conditions of intradimensional variability (2, 4, or 6 different stimulus values within each of the relevant and irrelevant dimensions). The precise nature of the stimulus dimensions and a verbal description of the stimulus values within each dimension under the different conditions of intradimensional variability are given in Table 1. Numerical measurements in this table refer to the size and position of the projected image on the screen in front of *S*.

As indicated in Table 1, *form* and *width* were the relevant dimensions for all conditions, so that the four response keys from left to right always corresponded to narrow triangles, narrow quadrangles, wide triangles, and wide quadrangles, respectively. Color was used as an irrelevant dimension for all conditions, position was irrelevant for conditions involving two or three irrelevant dimensions, and stripes were irrelevant only for conditions involving three irrelevant dimensions.

Intradimensional variability was manipulated either by increasing the number of stimulus values along the dimensions involved, as in the case of width and position, or by introducing other stimulus values on some different but related dimension, as with form, color, and stripes. In the former case, Table 1 shows that the stimulus values in the 2-value conditions were relatively similar although presumably readily discriminable by *S*, and that the added stimulus values for the 4- and 6-value conditions varied in the direction of a greater difference from stimuli of the other response category, thereby minimizing any effect of reduced perceptual discriminability between categories in the conditions of higher intradimensional variability. By using two different bases for introducing intradimensional variability, each of which was represented within relevant and irrelevant dimensions, it was hoped that the effects of this variable could be explored in a more general way than

would have been possible if only one method had been used.

Subjects and procedure.—The *S*s were 90 college students, each of whom was assigned unsystematically to one of nine equal-sized groups, representing all combinations of the three conditions each of intra- and interdimensional variability. All *S*s served individually and were presented at the outset with detailed instructions describing the nature of a concept, the operation of the controls, and the meaning of the signal lamps. The different dimensions and the possible values within each dimension were described in detail and illustrated by sample stimulus patterns. All *S*s were provided with slips of paper listing the five dimensions and the six possible stimulus values within each dimension, similar to the information contained in Table 1. These were placed next to the response panel and available to *S* for inspection throughout the experiment.

A separate stripfilm consisting of 256 patterns was prepared for each of the nine experimental conditions, but all were comparable insofar as possible with respect to the values taken by the relevant and irrelevant dimensions at corresponding positions within the series and the sequence of correct responses. In those conditions of low complexity where less than 256 different patterns were possible, an attempt was made to equate the number of times each pattern occurred. The experiment was terminated when *S* had either made 16 correct responses in succession, or had completed the entire sequence of 256 patterns without reaching this criterion.

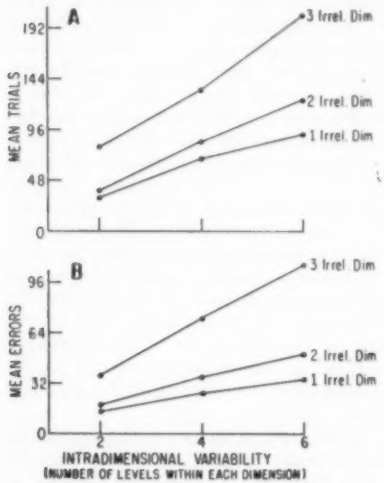

FIG. 1. Mean number of trials (A) and mean number of errors (B) as functions of intra- and interdimensional variability. Each plotted point represents the data from 10 *S*s.

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS AND TRIALS

Source	*df*	Errors *F*	Trials *F*
Interdimensional variability	2	18.80**	13.77**
Linear	1	33.60**	25.29**
Quadratic	1	3.92	2.26
Intradimensional variability	2	13.10**	18.72**
Linear	1	26.11**	37.44**
Quadratic	1	0.02	0.00
Interaction (Inter. × Intra.)	4	1.76	0.83
Linear × Linear	1	6.48*	3.03
Residual	3	0.19	0.09
Error	81	(972.1)	(3286.0)

Note.—Numbers in parentheses are error mean squares.

* $P = .05$.
** $P = .01$.

RESULTS

The results are summarized in Fig. 1, which presents both mean number of errors (Fig. 1B) and mean trial number on which the last error was made (Fig. 1A) for each of the nine experimental conditions. Both dependent variables show an increase as either intra- or interdimensional variability is increased, with a suggestion of a greater effect of intradimensional variability at the highest level of interdimensional variability, particularly for the errors measure.

Both measures were subjected to analysis of variance, summarized in Table 2, which indicated that the trends suggested by Fig. 1 are all highly significant. Increased interdimensional variability produces a consistent increment in both errors and trials, as shown by the highly significant linear component in each case. That the major part of the

change occurs between two and three irrelevant dimensions is suggested by the fact that the quadratic component approaches significance, particularly for the errors measure. The curve describing the effect of intradimensional variability is almost completely linear, since the linear component is not only highly significant but accounts for over 99.9% of the variance due to intradimensional variability in both measures.

The variance due to interdimensional variability is greater than that due to intradimensional variability for the errors measure, while the reverse is true for the trials measure. The only other difference between the two measures lies in the presence of a significant Linear × Linear component in the interaction for the errors measure, which is not significant for trials. This is attributable to the linear increase in slope of the intradimensional variability function with increasing interdimensional variability, as can be observed in Fig. 1B. No other interaction component approaches significance for either measure.

Curves presenting the mean number of errors for each block of 16 trials were also plotted, and showed that the effect of both variables remains highly consistent through practice. Because these curves indicated nothing essentially different from the overall measures, they have not been presented here.

Discussion

The effects of interdimensional variability correspond closely to the results of previous studies (Bourne, 1957; Bourne & Haygood, 1959), and indicate that the functional relationship observed in these studies also is found under the more complex conditions produced by increasing intradimensional variability.

The increase in errors with increasing intradimensional variability was found to correspond very closely to a straight line function, at least within the restricted range of conditions employed in this experiment. Because the methods of producing intradimensional variability were largely intuitive in nature, and differed somewhat for the various dimensions, further generalization beyond the present conditions must necessarily proceed with extreme caution. In order to get a rough indication of the effects of intradimensional variability on stimulus discriminability, absolute judgments were obtained from 24 *Ss*, most of whom had previously served in the experiment proper, for each of the five dimensions of 60 stimulus patterns used in the highest complexity condition (3 irrelevant dimensions, 6 values within each dimension). These results indicated that the dimensions did differ considerably in intradimensional discriminability, ranging from 7% errors on the stripes dimension to 33% errors on width. Insofar as these errors represented confusions between categories (e.g., between patterns to be responded to differently), the differences become critical for interpretation of the present results. The percentages of between-category errors for each dimension, given in the last column of Table 1, do not differ markedly for the various dimensions. The only sizeable discrepancy here concerns the stripes dimension, which had the fewest total errors, but ranked second only to width in between-category errors, 68% of its total being of this type. This suggests that the variation in number of stripes may have been more apparent than the horizontal-vertical distinction within this particular dimension. These unique characteristics of the stripes dimension may be partially responsible for the greater slope of the curves in Fig. 1 for three irrelevant dimensions, since stripes varied in this condition of interdimensional variability only.

In the light of the preceding discussion, any attempt to use the present data to extend or test a theoretical model is tenuous, and only a tentative suggestion as to the theoretical implications of the results can be justified. A decline in performance with increasing intradimen-

sional variability would be consistent however with any number of existing theoretical structures. For example, it could readily be accounted for in terms of a generalization framework (e.g., Gibson, 1940), or the concept of acquired distinctiveness and equivalence of cues (Dollard & Miller, 1950; Lawrence, 1949).

The present results are easily incorporated into the mathematical model of concept identification proposed by Bourne and Restle (1959), which states that the number of errors to solution is determined by the proportion of the total cues in the situation relevant to the solution of the problem. Since increasing intradimensional variability will have the principal effect of adding irrelevant cues, with a corresponding reduction in the proportion of relevant cues, an increase in number of errors would be predicted. Intradimensional variability would then be handled in essentially the same way as number of irrelevant dimensions was treated in the original paper (Bourne & Restle, 1959).

Assuming that the relationship between intradimensional variability and errors is linear, as the results strongly indicate, the Bourne-Restle model would predict that further increases in intradimensional variability beyond the range of the present experiment would soon result in a total inability of *Ss* to solve the problem at all. While such a prediction may seem implausible on the surface, it must be remembered that further increases in number of stimulus values would soon approach the informational capacity of the human organism (Miller, 1956), which might well preclude errorless performance. Observation of *Ss* under the more complex tasks of the present experiment also suggests that the ceiling of solvability of the problem may not be far beyond these conditions. In any case, a satisfactory fit to the present results can be produced through a number of introductions or modifications of parameters of the model. Since the present data are inadequate to select among these possibilities, no attempt will be made to present them here.

Summary

Concept identification, in which *Ss* learned to classify geometric patterns into one of four categories based on similarities in the dimensions of form and width, was investigated as a function of two methods of varying the complexity of the task. As in previous experiments, interdimensional variability was manipulated by including 1, 2, or 3 irrelevant dimensions of variation in each pattern. In addition, intradimensional variability was produced by presenting 2, 4, or 6 different stimulus values within each dimension. A 3×3 factorial design was used, with 10 *Ss* in each of nine groups.

A significant decline in performance was observed for increases in either intra- or interdimensional variability, with a suggestion of an interaction such that the effect of intradimensional variability increased at higher levels of interdimensional variability. The results were found to be compatible with the mathematical theory of concept identification of Bourne and Restle (1959), although a precise evaluation of the model was not possible under the present conditions.

References

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

Bourne, L. E., Jr. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

Bourne, L. E., Jr., & Haygood, R. C. The role of stimulus redundancy in concept identification. *J. exp. Psychol.*, 1959, **58**, 232–238.

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

(Received March 11, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 334–338

AN ALTERNATIVE CRITERION FOR THE ELIMINATION OF "VOLUNTARY" RESPONSES IN EYELID CONDITIONING [1]

THOMAS F. HARTMAN AND LEONARD E. ROSS

University of Wisconsin

A study by Spence and Ross (1959) demonstrated that the use of a latency criterion is effective in eliminating "voluntary" form responses from eyelid conditioning data. There is, however, no evidence concerning the usefulness of this criterion under different experimental conditions. It would appear that any variable affecting the latency distribution of voluntary form responses would also affect the usefulness of the latency criterion in eliminating these responses. Thus the latency criterion might well be inappropriate for studies employing experimental conditions differing from those of the Spence and Ross study.

The purpose of the present experiment was to investigate the applicability of latency criteria to data obtained under the experimental conditions typically employed in the Wisconsin laboratory, conditions that differ in several respects from those of the Spence and Ross experiment. A further purpose of the present study was to determine the usefulness of an alternative criterion for eliminating voluntary form responses.

METHOD

Subjects.—The *S*s were 12 men and 28 women from an introductory psychology course. One *S* who adapted to the UCS during the experiment was discarded.

Apparatus.—The apparatus used was essentially that described by Hartman and Grant (1960). The CS was an .8 mL. increase in brightness of a 10 cm. milk glass disk. Ambient room illumination was approximately 1.0 mL. CS duration was 700 msec. The UCS occurred 540 msec. following the onset of the CS and was of 200 msec. duration. The intensity of the puff was regulated by a 150-mm. column of mercury (2.9 psi).

Procedure.—All *S*s received 60 acquisition trials. The intertrial interval varied between 20 and 40 sec. with a mean of 30 sec. and the trials were programed by means of a Western Union tape transmitter. No ready signal was used. The recording equipment incorporated differentiators so as to give a recording of the first time derivative of the eyelid's movement. Two judges, one of whom had similarly classified responses in the Spence and Ross (1959) study, independently classified each response which preceded the UCS as being a conditioned response (C), a voluntary response (V), or questionable (?). The same form criteria that were used in the Spence and Ross study were used in assigning a response to one of the three categories. In terms of combined judgments there were six possible categories: VV, CC, V?, C?, ??, and VC. Latency and derivative measurements were made by another person.

RESULTS AND DISCUSSION

Figure 1 presents the latency distributions for responses judged as VVs, V?s, ??s, C?s, and CCs. When these distributions are compared with the Spence and Ross data it appears that a greater proportion of the voluntary form responses occurred at long latencies under the conditions of the present study. In this respect it should be noted that only

[1] This research was supported in part by the National Science Foundation and in part by the Research Committee of the Graduate School of the University of Wisconsin with funds from the Wisconsin Alumni Research Foundation. The authors would like to thank David A. Grant for the use of his equipment in the study.

31% of the VV responses of the present study fell in the 200–300 msec. interval while this interval contained 83% of the VV responses obtained in the Spence and Ross experiment. Also, while the Spence and Ross study found very few VV form responses from 400 msec. to the onset of the air puff there was a substantial number of VV form responses in this interval in the present experiment.

When the percentage of responses in each classification is compared with the similar percentage in the Spence and Ross study close agreement is found. For each classification the first percentage presented below is that of the present experiment, the second that found in the Spence and Ross study: VV's 29.9, 27.1; V?'s, 7.5, 8.6; ??'s 13.5, 5.5; C?'s 11.6, 13.9; CC's 36.9, 44.3; and CV's .6, .5.

The finding of more voluntary form responses at long latencies in the present study suggests that a latency criterion would probably be less satisfactory with these data than it was found to be in the Spence and Ross study. The procedure used to investigate this possibility was identical to that used by Spence and Ross to determine the effects of applying a latency criterion to their

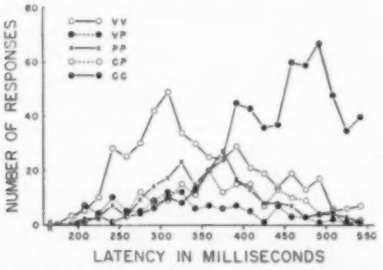

FIG. 1. Total number of responses in each judgment category as a function of response latency.

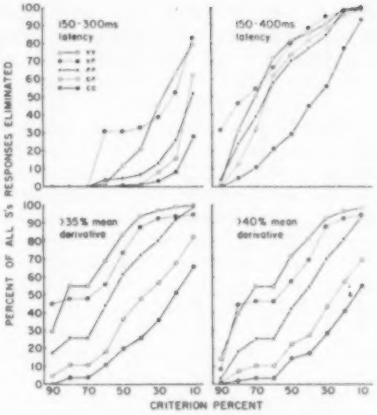

FIG. 2. Percentages of responses in each judgment category eliminated as a function of two latency and two derivative criteria used to define a voluntary responder.

data, and involved discarding *S*s using criteria from 90% to 10% in 10% steps. That is, *S*s were discarded if 90% or more of their responses met the latency criterion, 80% or more, etc., to 10% or more, and the percentage of responses in each category that would be eliminated for each of these criteria percentages was computed. Each graph thus presents the percentage of responses in each category eliminated as a function of the percentage of the total responses of an *S* that must meet the latency criterion (for that graph) in order for the *S* to be discarded.

When the resulting graphs (Fig. 2) are compared with the similar graph from the Spence and Ross study it is evident that latency criteria are less effective when used with the data of the present study. This is especially true of the comparable 150–300 msec. latency criterion.[2] While the 150–400

[2] A 150–300 msec. criterion was used in the present experiment rather than 200–300 msec. since Spence and Ross found some voluntary

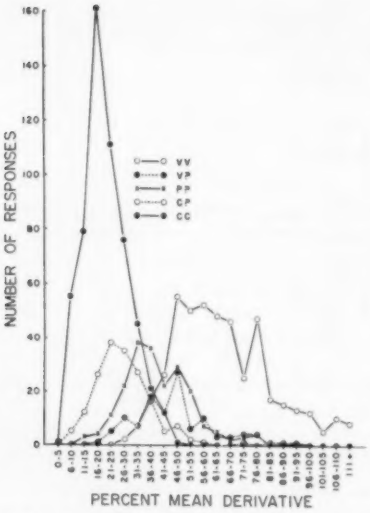

FIG. 3. Total number of responses in each judgment category as a function of percent of mean UCR derivative amplitude.

msec. criterion is more satisfactory, it results in the elimination of large percentages of ?? and C? responses which is probably not desirable.

Since one of the principle form criteria used to identify a response as voluntary is a sharp smooth closure, it was decided to determine the feasibility of using the slope of the closure as the criterion, independently of the latency of the response. Since the eyelid apparatus used gave a derivative record as well as the actual response form record, an objective measure of the slope of each response was available in terms of the amplitude of the derivative record for the response.

Accordingly, the derivative amplitude of each response was compared to the mean derivative amplitude obtained for the *S* on the first five unconditioned responses to the UCS, or available fraction thereof. In Fig. 3 the number of responses in the categories are shown as a function of the percentage of the mean UCR derivative amplitude. From the graph it appears that a cutoff of approximately 40% should be most effective in eliminating V form responses. Curves showing the percentage of VV, V?, ??, C? and CC, responses eliminated by derivative criteria of 40% and 35% are presented in Fig. 2. Similar to the procedure used for latency criteria, any response whose derivative amplitude was equal to or greater than the criterion percentage was counted as a voluntary response and the graphs show, as in the latency criteria graphs, the percentage of the various classes of responses eliminated as a function of the percentage of the total responses of an *S* that must meet the derivative criterion for the data for that *S* to be discarded. A 45% derivative criterion was also tried but since the results were essentialy the same as with the 40% criterion the graph is not shown.

It is evident that the use of a derivative (slope) criterion is more effective than a latency criterion under the experimental conditions of the present study, and is to be preferred under these conditions if it is considered desirable to eliminate voluntary-form responses from the data. It should be noted that a 150–400 msec. latency criterion eliminates a considerably greater percentage of responses judged V?, ??, and C? than are eliminated by the derivative criteria. Both the 35% and 40% derivative criteria appear to result in a satisfactory elimination of voluntary responses at a reasonable "cost" of C form responses. The use of a 35% derivative with a 40% or 50% *S* criterion

responses in the 180–200 msec. interval and therefore it appeared that 150–300 msec. might be more effective. A 150–350 msec. latency criterion was also tried, but did not prove as satisfactory as the 150–400 msec. criterion presented.

would appear to be a reasonable compromise. If a more or a less severe elimination is desired it can easily be obtained by the choice of the criterion percentage used to identify *Ss* as voluntary.

From observing *Ss* in the conditioning situation under ready signal and no ready signal conditions, as well as examining the records of a few *Ss* instructed to blink with and without a ready signal, it appeared likely that, due to attentional factors, the presence or absence of a ready signal was a major factor in determining the latency of voluntary responses. Since the experimental conditions of the present study differ from those of the Spence and Ross experiment in a number of aspects besides the ready signal-no ready signal variable (i.e., number of trials, CS-UCS interval, CS and UCS duration, *Ss* dark adaptation conditions, and intertrial interval), the longer latencies of the present study cannot be attributed to a single factor on the basis of the data from this experiment. However in an effort to help determine to some degree the relationship of the different variables to voluntary response latency, the voluntary-form response latency data from several studies were analyzed. Thus the voluntary-form latency plots of the first and second halves of a study by Beeman, Hartman, and Grant (1960) were compared and no apparent differences found; unpublished data (Boyes & Grant) revealed no relationship between the number of *Ss* classified as voluntary (on the basis of response form) and five CS durations; and other data (Hansche, 1959) indicated that with a 700 msec. CS-UCS interval the distribution of responses from *Ss* instructed to make a light go on (accomplished by blinking) peak at about the same point as the VVs in the present study. Latency distributions of responses from other *Ss* instructed to blink to the onset of the CS have been found to be quite similar under dark adapted and high illumination conditions, and while the data of some *Ss* under these instructions do indicate that intertrial interval might have an effect under no-ready-signal conditions, the data do not suggest that intertrial interval was an important factor in producing the voluntary response latency differences being considered. It also should be pointed out that the Iowa laboratory has found the latency criterion to be satisfactory with a wide range of puff intensities and puff durations with both men and women, indicating that the latency of voluntary responses has been found to be relatively invariant under these conditions.

Although the analyses mentioned above were not systematic and small numbers of *Ss* were involved in some cases, the results tend to support the suggestion that the ready signal-no ready signal variable is of primary importance in producing differences in the voluntary-form latency distributions. Also, if the voluntary response can be regarded as essentially similar to reaction time responses we might expect the voluntary response latency to be a function of variables that have been found to affect simple reaction time. In this regard it is interesting that the foreperiod has been demonstrated to be an important variable in reaction time studies (Woodworth, 1938, p. 314), and if we consider the intertrial interval to be the foreperiod under no-ready-signal conditions and the ready signal-CS onset period to be the foreperiod when a ready signal is employed, the ready-signal conditions should result in shorter reaction times, i.e., shorter response latencies, than would be the case with the more variable and longer (less optimal) "foreperiods" of the no ready signal condition.

It would seem unwise in view of the data obtained in the present study to apply a voluntary response criterion to data collected under experimental conditions varying to a marked degree from those for which a criterion has been demonstrated to be applicable. The practice of using a latency criterion when no ready signal is employed would appear especially questionable.

Summary

The eyelid conditioning data of 40 *Ss* were analyzed in order to investigate the usefulness

of latency criteria in eliminating voluntary form responses from eyelid conditioning data obtained under experimental conditions differing from those of a study which demonstrated the usefulness of a latency criterion.

The results indicated that latency criteria were inappropriate under the experimental conditions of the present study due to the distribution of voluntary response latencies, but that a criterion involving the slope of closure (time derivative amplitude) was a satisfactory substitute.

It was suggested that the principal factor producing the differences in the voluntary response latency distributions of the two studies was the presence or absence of a ready signal preceding each trial. The present study did not employ such a warning signal.

References

BEEMAN, E. Y., HARTMAN, T. F., & GRANT, D. A. Supplementary report: Influence of intertrial interval during extinction on spontaneous recovery of conditioned eyelid responses. *J. exp. Psychol.*, 1960, **59**, 279–280.

HANSCHE, W. J. A comparison of operant, avoidance, and classical technique in the conditioning of the eyelid response. Unpublished doctoral dissertation, University of Wisconsin, 1959.

HARTMAN, T. F., & GRANT, D. A. Effect of intermittent reinforcement on acquisition, extinction, and spontaneous recovery of the conditioned eyelid response. *J. exp. Psychol.*, 1960, **60**, 89–96.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 17, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 339–344

SECONDARY REINFORCEMENT BASED ON STIMULUS-CHANGE PRIMARY REINFORCEMENT

CARL L. ROBERTS, KENNETH E. LEBOW[1], AND ROBERT M. YODER

Colorado College

In recent years many investigators have been demonstrating that stimulus changes not clearly correlated with drive operations and not previously paired with drive-correlated reinforcers have at least two of the properties of primary reinforcers. Their presentation increases the strength of preceding responses and withholding them produces extinction. Seward (1942), for example, removed *Ss* from a response chamber following a response and Lawson and Dawson (1957) and Gentry (1958) removed the chamber from around *Ss*. All studies produced evidence of learning of the response preceding these operations although the learning was not as strong as when food pellets followed the response (e.g., Lawson & Dawson, 1957). Other workers have used such stimulus events as light onset (Marx, Henderson, & Roberts, 1955; Roberts, Marx, & Collier, 1958), click onset (Barnes & Kish, 1961), and a view of a monkey colony (Butler, 1953), with the same results. In spite of the lack of systematic experimental comparisons of strength between these types of reinforcers and food and water for a given species, some writers have referred to the former examples as low-efficacy primary reinforcers (e.g., Kling, Horowitz, & Delhagen, 1956).

The question arises whether any such weak reinforcers have a third property traditionally attributed to primary reinforcers, viz., that on their basis one may make of a so-called "neutral" stimulus a secondary reinforcer (S^r), by pairing it with a known primary reinforcer in appropriate ways.

The primary purpose of the present experiments was to determine whether such stimulus-change primary reinforcers as mentioned above could perform this S^r-making function. The fact that such stimuli as light onsets and clicks have now been shown to have "intrinsic" reinforcing properties alters neither the probability that what basic reinforcing power they do have may be enhanced by pairing them with presumably more powerful primary reinforcers such as food and water, nor the possibility that a secondary reinforcer is typically a weak primary reinforcer whose effectiveness has been enhanced through pairing with a stimulus which has a more powerful reinforcing effect.

An efficacy scale of reinforcers is as yet lacking. Therefore, for present purposes, we supposed that the gross and abrupt removal of a response box from around *Ss* as shown by Lawson and Dawson (1957) and Gentry (1958) is a more effective reinforcer of a preceding response than a simple increase in illumination (light onset), and that food pellets are in turn more effective reinforcers than box removals. Consequently, in the present experiments, we predicted that if we paired light onsets with box removals according to typical, magazine-training procedures as used in the establishment of S^r's, then light onset with such a history would subsequently reinforce an operant more effectively than light onset without such a history. Finally,

[1] Now at the University of Massachusetts.

we predicted that light onsets paired with food pellets would become a more powerful S^r than light onsets paired with box removals.

Experiment I[2]

Method

Subjects.—The *Ss* were 40 naive male albino rats, aged 150–180 days at the start of experimentation. Each *S* lived throughout the experiment in a 9½ × 7 × 7 in. Wahman rat cage. The *Ss* were fed 10–12 gm. of Purina lab chow in pellet form 30 min. after the end of daily trials but had free access to water at all times except while in the experimental apparatus.

Apparatus.—The basic apparatus consisted of four boxes, 12½ in. high and 8½ in. square. These boxes were painted black on the inside and had no bottoms. Each rested in a larger (11½ in. high and 12 in. square), unpainted, retaining box. All boxes were made of ¼-in. plywood. A cross bar was bolted jointly to the projecting tops of three of the bottomless boxes. To the center of this an eyebolt was attached such that a rope running through ceiling pulleys could be pulled to jerk the smaller boxes out of the larger ones simultaneously. The fourth box had a food cup of ½-in. diameter inside, 2¼ in. from its bottom. From this cup a tube, through which food pellets could be dropped, ran outside the box. All boxes had a ¾-in. wide and ¾-in. high vertical slot in one side beginning 3 in. from the bottom of the box (1½ in. from the food cup in the fourth box), through which a lever could be inserted. The lever mechanisms were mounted on the outsides of the bottomless chambers and the lever crosspieces were slipped into a bearing through the slot when desired. Lever presses were recorded by Veeder-Root electrical counters housed in another room. A 7½-w. frosted bulb penetrated partly into each box 7 in. above the lever. Across from the bulbs were 7½-in. × 4½-in. sheets of aluminum which reflected the lights. The lights could be turned on either by a master foot switch or individually by a lever press in each box; they remained on until manually switched off or the levers were released.

Procedure.—The procedure was as follows: The rats in Group I, a traditional-type, S^r-training group, received light-onset presentations followed within 2 sec. by the presentation of a Purina Rabbit Chow Checker. The light went off upon the arrival of the pellet in the food cup. (The *Ss* were given pretraining during which they learned to come to the food cup and eat when the pellet hit the cup.) The procedure involved a modification by Gentry (1958) of a technique suggested by Zimmerman (1957). The *Ss* were given a total of 344 light onsets but these were followed by a total of only 102 food pellets. Specifically, for the first 16 trials at 4 trials a day, each light onset was followed within 2 sec. by a food pellet. After *Ss* had completed these 16 trials, the mean ratio of food pellets to light onsets for the next 16 trials was decreased to one food pellet after every second light onset. This means that, randomly, one-half of the 16 trials were at this ratio while one-fourth were at a 1:3 ratio and one-fourth at a 1:1 ratio of food pellets to light onsets. After a series of 16 trials of this order, the mean ratio was decreased to one food pellet after every third light onset—and so on until *Ss* were getting one food pellet after every sixth light onset. Light onsets were approximately 30 sec. apart.

Groups II and III were treated exactly as Group I except that box removals from around *Ss*, rather than food pellets, followed the light onsets. Group IV was treated exactly as Groups II and III except that no light onsets preceded box removals. Also, for Groups II, III, and IV, a wire screen was slid over the retaining boxes after the bottomless chambers were removed and hanging above the retaining chamber. All *Ss* remained in the retaining chambers for 30 sec. after the box removal.

One *S* from each group was run at the same time each day to keep food-privation level constant and the 4-trial blocks were spaced at 24-hr. intervals. Testing for the effect of light onset as an S^r began 24 hr. after the last training session. The *Ss* were given seven 5-min. lever-pressing sessions at 24-hr. intervals. In these sessions lever pressing was followed by light onset for the duration of each lever press for Groups I, II, and IV. Group III received no light onsets for lever pressing.

Results and Discussion[3]

The data of the test sessions, plotted as mean square root number

[2] The authors wish to acknowledge the careful assistance of the late Sharon Westerfeld in the collection of these data.

[3] A square-root transformation was used to eliminate heterogeneity of variance in both experiments here reported. All analyses were performed on square-root values.

of responses, are shown in Fig. 1. A Treatments × Sessions analysis of variance of number of lever presses, presented in Table 1, shows the main effects of both treatments and sessions, as well as their interaction, to be significant sources of variance. It is clear from the curve of Group II that light onset, which had been discriminatively paired with box removals, reinforced subsequent lever pressing. That the light onset paired with food pellets had acquired such a capacity only slightly is indicated by the curve for Group I. This group is differentiable from Group IV, for which light onset had served as a low-efficacy primary reinforcer (cf. Kling et al., 1956; Roberts et al., 1958) only during Sessions 1, 2, and 3. The interaction resulted, in fact, from the intermingling of these two groups over sessions. Finally, there is the over-all downward trend typical of the operant level (Group III), and of responses under the control of S^r's (Group II).

There would seem to be at least four possible explanations of why *Ss* which had had light onsets paired with box removals were so clearly reinforced by the light onsets, whereas the food-pellet group showed little, if any, such secondary reinforcement. First, the light did not overlap with the beginning of food-pellet consumption by the rats, whereas, as a result of procedural technique, it did often overlap with the initial movements of box removal. As soon as the pellet hit the food cup the light went off. Therefore, food pellets may have often been paired with light offset as well as with light onset. Second, there is a remote possibility that just getting lights turned on a great number of times enhances light onset's reinforcing power—as is the case for rats under certain other illumination conditions (Ochocki & Premack, 1958; Roberts, 1959; Roberts et al., 1958)—and that it was this, rather than the associations of light with box removal, that produced the reinforcing effect of light onset. Why Group I should not then have behaved in the same way would remain, of course, unexplained. Third, light onset was, for the food group, a discriminative stimulus (S^D) "once removed," where, for the box-removal group it served as the only S^D, watching-the-box-go-up being the apparent response in this case. That is, *Ss* in Group I first learned to approach the food cup upon the click of the entering food pellet. Light onset was then allowed to precede the click and was, hence, less closely related in time to the food pellet. Fourth, during lever pressing, the response chain initiated by the light $S^{r \cdot D}$ for food approach was possibly much longer than the chain initiated by light onset in the box-removal group. In short, going to the food cup, sniffing around, etc., takes longer than just looking up and about, the apparent responses initiated by light

TABLE 1

ANALYSIS OF VARIANCE OF SQUARE-ROOT NUMBER OF RESPONSES: EXP. I

Source	*df*	*MS*	*F*
Between *Ss*	39		
Treatments (T)	3	20.725	8.58**
Error (b)	36	2.416	
Within *Ss*	240		
Sessions (S)	6	1.665	3.15*
T × S	18	1.059	2.01*
Error (w)	216	0.528	
Total	279		

* $P < .01$.
** $P < .001$.

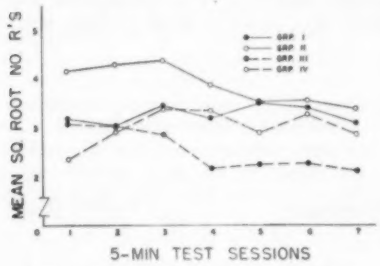

FIG. 1. Response frequency over seven 5-min. test sessions.

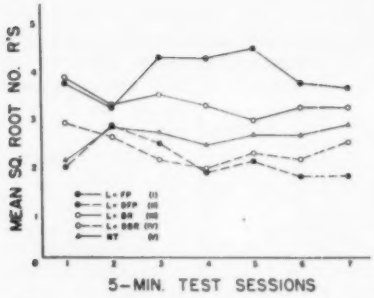

FIG. 2. Response frequency over seven 5-min. test sessions. Letters refer to training conditions: L means light; FP, food pellet. DFP, delayed food pellet; BR, box removal; DBR, delayed box removal; NT, no training;

onset when box removal has been the primary reinforcer of these responses.

To check possibilities 1 and 2 a second experiment was performed.

EXPERIMENT II[4]

Method

Subjects and apparatus.—Fifteen naive male, and 8 naive female, albino rats, between 170 and 200 days old, served as *S*s under the same maintenance conditions as in Exp. I. The apparatus was the same as that used in Exp. I.

Experimental design and procedure.—Group I ($N = 3$ males, 2 females), received the same treatment as in Exp. I except that the light stayed on for ½ sec. after the food pellet hit the food cup. Group II ($N = 3$ males, 2 females), was the control for Group I. It was treated identically to Group I except that 1 min. intervened between the onset of light and the presentation of the food pellet. Group III ($N = 3$ males, 2 females), received light onsets paired with *box removals* and the ratio of light-onset—box-removal pairings was exactly as in the case of Group I. The procedure here was that when a light onset was to be followed by box removal, the light came on and within 2 sec. the bottomless box was removed from around *S* via the overhead pulleys, leaving *S* in the retaining chamber. The light went off ½ sec. after the box went into the air and a screen was slipped over the top of the retaining chamber. Group IV ($N = 3$ males, 1 female), the control for Group III, was treated like Group III except that 1 min. intervened between light onset and box removal. As in Exp. I all *S*s stayed in the retaining chambers for 30 sec. after the presentation of food pellets or box removals. Group V ($N = 3$ males, 1 female), received no training. Its members simply remained in their home cages during training and subsequently provided a measure of the intrinsic reinforcing effect of light onset alone.

Following the 24-day training phase, all *S*s were given seven 5-min. lever pressing sessions at 24-hr. intervals with all presses followed by light onset. One *S* from Groups I–IV was run at the same time each day throughout experimentation to keep food-privation level constant across groups. Group V was run last and as a group.

Results

The data of the test sessions are shown in Fig. 2. An analysis of variance of number of lever presses, summarized in Table 2, shows treatments to be a significant source of variance. The curve of Group I, the food-pellet group, shows clearly that light onset had acquired the power to reinforce, and the curve for Group III, the box removal group, points strongly in the same direction in that it is well above the curve of Group V, for which light onset served as a primary reinforcer. Further, the difference between these two groups is in the direction one might expect on the *assumption* that food was in some sense a more effec-

TABLE 2

ANALYSIS OF VARIANCE OF SQUARE-ROOT NUMBER OF RESPONSES: EXP. II

Source	*df*	*MS*	*F*
Treatments	4	17.827	13.26**
Within treatments	18	1.344	
Total	22		

** $P < .001$.

[4] A brief report of this experiment with allusions to Exp. I was made at the meetings of the Midwestern Psychological Association, Chicago, 1959.

tive S^r-maker because a more effective primary reinforcer than box removal. Differing findings on whether the power of an S^r is a function of amount of primary reinforcement under non-comparison conditions suggest that whether it is may depend on the R-class and on the type of primary reinforcement (cf. Butter & Thomas, 1958; Lawson, 1953; Stebbins, 1959).

These findings would seem to support Explanation 1 of the findings of Exp. I, negate Explanation 2, and make Explanations 3 and 4 highly improbable.

Discussion

The results of these two experiments provide evidence that a stimulus can acquire a reinforcing capacity which is greater than its intrinsic reinforcing power by being paired with another stimulus with no apparent drive-relevancy such as food and water have. Further, that such secondary reinforcement can be produced in this way suggests that any stimulus changes which can do this S^r-making work should be regarded as "primary" reinforcers and, finally, that a secondary reinforcer is simply a strengthened "weak" primary reinforcer. Of what sort and how "large" stimuli will have to be to serve as S^r-making primary reinforcers is an empirical question. Perhaps any stimulus which produces or controls a higher rate or greater magnitude of response than another stimulus can make of the second stimulus an S^r, which may mean simply the enhancement of an already existent reinforcing effect (cf. Premack, 1959).

The problem of whether to regard such stimulus changes as box removal as relevant to some drive, e.g., "curiosity," awaits the clear operational definition of such constructs. Interestingly enough, previous attempts to relate such reinforcement to deprivation for stimulation (Ochocki & Premack, 1958; Roberts, 1959; Roberts et al., 1958) showed that rats maintained in "stimulus rich" environments responded more to stimulus-change reinforcement than did rats maintained in "stimulus-poor" surroundings. There is at least one study (Butler, 1957), however, that gives credence to the view that stimulus-change reinforcement may be an increasing function of stimulus deprivation.

The present findings are also relevant to the question of whether a stimulus has to be an S^D before it can become an S^r (cf. Schoenfeld, Antonitis, & Bersh, 1950). If so, then for what Rs is light onset an S^D when light onset is paired with box removal? Observations made during the present experiments indicate that these Rs are "looking-up-and-about." Careful observation is required in this case as well as in other stimulus-change reinforcement research to determine the nature of the various "consummatory" behaviors.

Finally, it is interesting that the responses showed little, if any, tendency to extinguish, particularly in Exp. II. We considered, of course, that we were dealing with a subtle or labile phenomenon. Consequently, we used very short test sessions with long intersession intervals and a modification of Zimmerman's method (1957) to establish durable secondary reinforcement. On the face of it these devices would appear to have been successful.

Summary

Two experiments were carried out to determine whether a light onset discriminatively paired with the removal of a box from around *Ss* would become a secondary reinforcer. It did. The findings were briefly discussed in relation to the nature of secondary reinforcement, of such primary reinforcement as "box removal," and to relevant drive operations. Also, the question was raised as to what are the Rs for which light onset is the discriminative stimulus, when light onset is paired with box removal.

References

Barnes, G. W., & Kish, G. B. Reinforcing properties of the onset of auditory stimulation. *J. exp. Psychol.*, 1961, **62**, in press.

Butler, R. A. Discrimination learning by rhesus monkeys to visual-exploration mo-

tivation. *J. comp. physiol. Psychol.*, 1953, **46,** 95–98.

BUTLER, R. A. The effect of deprivation of visual incentives on visual exploration motivation in monkeys. *J. comp. physiol. Psychol.*, 1957, **50,** 177–179.

BUTTER, C. M., & THOMAS, D. R. Secondary reinforcement as a function of the amount of primary reinforcement. *J. comp. physiol. Psychol.*, 1958, **51,** 342–348.

GENTRY, T. A. Secondary reinforcement on the basis of non-reward primary reinforcement. Unpublished master's thesis, Colorado College, 1958.

KLING, J. W., HOROWITZ, L., & DELHAGEN, J. E. Light as a positive reinforcer for rat responding. *Psychol. Rep.*, 1956, **2,** 337–340.

LAWSON, R. Amount of primary reward and strength of secondary reward. *J. exp. Psychol.*, 1953, **46,** 183–187.

LAWSON, R. Brightness discrimination performance and secondary reward strength as a function of primary reward amounts. *J. comp. physiol. Psychol.*, 1957, **50,** 35–39.

LAWSON, R., & DAWSON, D. R. A refinement of Seward's test of Guthrie's theory. Paper read at Midwestern Psychological Association, Chicago, May 1957.

MARX, M. H., HENDERSON, R. L., & ROBERTS, C. L. Positive reinforcement of the bar-pressing response by a light stimulus following dark operant pretests with no aftereffect. *J. comp. physiol. Psychol.*, 1955, **48,** 73–76.

OCHOCKI, F. J., & PREMACK, D. The joint effect of stimulus deprivation and intertest interval on the frequency of light-contingent bar pressing. Paper read at Midwestern Psychological Association, Detroit, May 1958.

PREMACK, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychol. Rev.*, 1959, **66,** 219–233.

ROBERTS, C. L. Stimulus and stimulus-change factors governing the free operant rate. Unpublished doctoral dissertation, University of Missouri, 1959.

ROBERTS, C. L., MARX, M. H., & COLLIER, G. Light onset and light offset as reinforcers for the albino rat. *J. comp. physiol. Psychol.*, 1958, **51,** 575–579.

SCHOENFELD, W. N., ANTONITIS, J. J., & BERSH, P. J. A preliminary study of the training conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1950, **40,** 40–45.

SEWARD, J. P. An experimental test of Guthrie's theory. *J. exp. Psychol.*, 1942, **30,** 247–256.

STEBBINS, W. C. Relation of amount of primary reinforcement to discrimination and to secondary reinforcement strength. *J. comp. physiol. Psychol.*, 1959, **52,** 721–726.

ZIMMERMAN, D. W. Durable secondary reinforcement: Method and theory. *Psychol. Rev.*, 1957, **64,** 373–383.

(Received March 17, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 345–349

THE ROLE OF OVERLEARNING TRIALS IN DETERMINING RESISTANCE TO EXTINCTION[1]

NATHAN R. MURILLO[2] AND E. J. CAPALDI

University of Texas

Reid (1953) has shown that overlearning trials facilitate response reversal in a discrimination learning situation. He has interpreted this result in terms of the growth of a mediating reaction termed the re-response of discriminating. This response is hypothesized to occur late in training and to gain strength only after the overt response (e.g., left turning) has itself approached asymptotic or near asymptotic strength. In sum, *S* learns to respond to a subset of the total stimulus situation provided by the environment. This formulation serves as the point of departure for the present investigation, the purpose of which is to determine the effects of overlearning trials on resistance to extinction. Extinction, of course, can probably best be considered as a variety of transfer of training.

Several studies (Capaldi, 1957, 1958; Lewis & Duncan, 1956, 1958; Stimmel & North, 1959) have demonstrated an inverse relation between amount of training and resistance to extinction. The hypothesis adopted in this investigation is that extinction is reduced following increased training if—and only if—such increased training includes overlearning trials. A regular pattern of events was utilized in the attempt to test the hypothesis. A regular pattern allows the precise specification of the number of trials required to meet the criterion of learning and also of the number of overlearning trials involved.

[1] This research was supported by a grant from the University of Texas Graduate School.

[2] Now at the University of Houston.

METHOD

Subjects.—The *S*s were 88 undergraduate men and women volunteers from the introductory psychology course at the University of Texas. The *S*s were assigned to one of four groups on the basis of order of appearance at the experimental place.

Apparatus.—The apparatus has been described elsewhere (Capaldi, 1957). It can be described as a modification of the Wisconsin General Test Apparatus (Harlow, 1951). It consisted of a partition which obscured *S*'s view of *E*, a movable tray which could be pushed toward or drawn away from *S*, with a single well in the tray covered by a lid. A guillotine door placed in the center of the partition, when raised, permitted the tray to be freely drawn; when closed, it permitted *E* to bait the well without *S*'s knowledge. The well was baited with a piece of green velvet cloth. The contact of the cloth with the well was noiseless. The entire apparatus was painted gray.

Instructions.—The *S* was seated before the apparatus and was given the following instructions:

> When this door is raised, this tray can be pushed toward you. This lid conceals a little well which you can see now that I have lifted the lid. You are to tell me if you think this (*E* shows *S* the piece of green velvet cloth) is under the lid when the tray is pushed toward you. Make the best guess you can. If your guess is that the cloth is under the lid, say 'in'; if your guess is that the cloth is not under the lid, say 'out.' After having given your guess, lift the lid to determine if you were correct. Then replace the lid. Remember, I want the best guess you can make under the circumstances. Do you understand?

It should be noticed that the tray was baited and pushed toward *S* prior to the time a guess was given in an attempt to prevent *S* from considering that the character of the

trials, i.e., positive (cloth in well) or negative (absence of cloth in well) was determined by the response given.

Design.—The regular pattern of events used demanded that every third trial be positive, the preceding two trials negative. The four groups of *Ss* received either 12, 24, 48, or 60 acquisition trials. Immediately following the last acquisition trial, extinction training was begun; i.e., all trials were negative. The criterion of extinction was eight consecutive negative responses. If during acquisition training *S* had given seven consecutive correct responses, he was considered a "learner." All other *Ss* were considered to be "nonlearners." Whether or not *S* learned had no effect upon the manner in which acquisition or extinction training was given. In short, the appropriate numbers of acquisition and extinction trials as indicated above were given regardless of the character of *S*'s responses. A few remarks are in order concerning the choice of experimental design.

Pilot investigation indicated that approximately 40% of all *Ss* failed to learn the pattern even after inordinate amounts of training. On the other hand, approximately 30% or so of *Ss* proved capable of learning the pattern in very few trials (4, 5, or 6). These results are consistent with the usual findings of studies dealing with pattern learning (Goodnow & Pettigrew, 1956).

It is clear that the peculiarities of pattern learning create difficulties for the usual design dealing with overlearning; i.e., *Ss* are matched on the basis of speed of original learning and different groups are given a greater or lesser number of overlearning trials. However, it was soon realized that the failure of some *Ss* to learn might not necessarily be disadvantageous, but rather this fact might be turned to scientific advantage. Certain experimental facts which will be discussed later suggest the correctness of the above assumption. The major advantage of the design is that it allows comparisons between (*a*) nonlearners who have received different amounts of training, (*b*) learners who have received different numbers of overlearning trials, and (*c*) learners and nonlearners who have received the same amount of training. It should be realized that the present design depended upon the ability to predict with some degree of accuracy the number of learners to be expected under a given amount of training.

TABLE 1

NUMBER OF LEARNERS, MEAN TRIALS TO CRITERION, AND MEAN NUMBER OF OVERLEARNING TRIALS IN EACH GROUP

Group	*N*	Mean Trials to Criterion	Mean Overlearning Trials
12	5	4.40	0.60
24	9	8.33	8.66
48	11	14.18	26.82
60	13	14.23	38.61

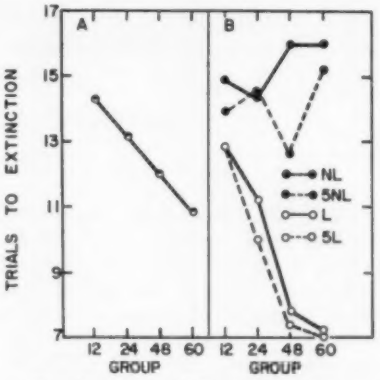

FIG. 1. Mean trials to extinction as a function of amount of training. (A) All *Ss* in each group. (B) The *Ss* who met (L) or failed to meet (NL) the criterion of learning. Solid lines represent all learners or nonlearners; broken lines, a random selection of five cases from each group, except the 12-trial group of learners.

RESULTS

Pattern learning.—The majority of *Ss* who learned the pattern did so within 10 trials or less. In the 60-trial group 8 *Ss* learned within 10 trials or less, 5 *Ss* were distributed in the range from 11 to 47 trials, and 9 *Ss* failed to learn the pattern. The corresponding figures for the 48-trial group were 7, 4, and 11 *Ss*, respectively.

In Table 1 are presented the data on the number of learners in each of the groups, the mean number of trials required by these *Ss* to begin meeting the criterion of learning, and the mean number of overlearning trials received by each group of learners. The learners in the 12-trial group received

virtually no overlearning trials. The number of overlearning trials received increases, of course, as amount of training does and reaches a rather high value in the 60-trial group.

Resistance to extinction.—The mean number of trials required by each of the groups to meet the criterion of extinction is shown in Fig. 1. As Section A of Fig. 1 shows, as amount of training increased, resistance to extinction decreased. This trend, although not significant ($F < 1$), is consistent with the results of earlier studies mentioned previously.

Section B of Fig. 1 shows a breakdown of the data presented in Section A. Resistance to extinction is shown as a function of meeting or failing to meet the criterion of learning. The solid lines represent a plot of all the data and the broken ones are plots resulting from a random selection of five cases from each of the groups of learners and nonlearners except the 12-trial group of learners. The purpose of the random selection of cases will be discussed below. As can be seen the nonlearners were considerably more resistant to extinction than the learners. Further, each group of nonlearners has a higher mean and each group of learners a lower mean than the corresponding group mean shown in Section A of Fig. 1. Little consistent trend can be found in the extinction of the nonlearners. However, the learners show a marked tendency toward decreased resistance to extinction with increased training.

The random selection of cases mentioned above was made because of the presence of disproportionality in the subclasses. The number selected from each subclass was dictated by the presence of five learners in the 12-trial group. The procedure of random selection of cases results in a weak test of significance (Wert, Neidt, & Ahmann, 1954). Nevertheless, it was used for two reasons. First, the differences shown in Section B of Fig. 1 are rather large. Secondly, since separate tests employing all the data were to be conducted as a check, no loss of information was foreseen. The analysis indicated that amount of training was not significant ($F < 1$). However, Learning ($F = 40.23$, $df = 1/32$) and the interaction between Learning and Amount of Training ($F = 3.93$, $df = 3/32$) were significant beyond the .01 and .05 levels, respectively. The separate tests are reported below.

Comparisons between learners and nonlearners at each of the values of training indicated that only the highly trained groups differed significantly ($t = 3.41$ for 48-trial group comparison and $t = 2.58$ for 60-trial group comparison, $P < .02$ for 20 df in each case). The lack of significant differences at the lower values of training and their presence at higher values is consistent with the significant interaction reported above.

Analysis of variance indicated that differences among the means of the nonlearners are not significant ($F<1$) whereas the differences among the means of the learners are significant beyond the .05 level ($F = 2.78$, $df = 4/33$). These analyses are also consistent with the one reported earlier.

The above analyses strongly support the hypothesis that overlearning trials are responsible for reduced resistance to extinction but in order to leave no doubt it is necessary to correct for the unequal rates of learning among the groups shown in Table 1. Two statistical means of control are at our disposal. Both are reported below.

If there is a relationship between Resistance to Extinction and Overlearning trials, then it is to be expected that when the Trials to Cri-

TABLE 2

NUMBER OF RAPID LEARNERS IN EACH GROUP, MEAN NUMBER OF TRIALS TO THE CRITERIA OF LEARNING AND EXTINCTION, AND MEAN NUMBER OF OVERLEARNING TRIALS

Group	*N*	Trials to Criterion	Trials to Extinction	Overlearning Trials
12	5	4.40	12.80	0.60
24	6	4.66	9.66	12.33
48	7	5.28	6.43	35.86
60	6	5.16	7.00	47.83

terion variable is partialed out the correlation between the two above mentioned variables will be significant. Resistance to Extinction is highly correlated with Trials Overtrained ($r = -.540$, $df = 37$, $P < .01$) but shows only a slight nonsignificant relationship with Trials to Criterion ($r = .039$). By partialing out Trials to Criterion the correlation of $-.540$ increases to $-.568$ which is significant beyond the .01 level. Another correlation will prove of interest in a short time. The r of .039 becomes $-.22$ when overlearning trials are partialed out. This correlation is not significant.

A second statistical test is at our disposal. In Table 2 are presented the data on rapid learners, i.e., those *S*s who met the criterion of learning in 6 trials or less. Trials overtrained, of course, vary, being greater in the more highly trained groups. An analysis of variance applied to these data yields an F significant well beyond the .01 level ($F = 13.93$, $df = 3/20$).

DISCUSSION

The major finding of the present experiment, and indeed the finding from which the remainder of the results can be deduced, was that resistance to extinction is decreased following increased training only if such training included overlearning trials. This finding supports the experimental hypothesis. Several consequences follow from the hypothesis.

One implication is that the mere learning of a pattern of reinforcement or other relevant response is not sufficient to result in reduced resistance to extinction as had been claimed (Capaldi, 1958). The present results bear out this deduction. Unless overtraining was involved, learners failed to extinguish more rapidly than nonlearners. The critical factor or factors responsible for reduced resistance to extinction seem to occur in the overlearning trials as the formulation of Reid (1953) suggests.

A second but related implication is that increased training per se without the occurrence of prior learning will not result in decreased resistance to extinction. The failure of the nonlearners to differ among themselves with increased training, whereas the learners did show a significant tendency toward decreased resistance to extinction, supports the above generalization. This generalization may aid in demonstrating the empirical and theoretical continuity of earlier findings (Perin, 1942; Williams, 1938) with more recent ones dealing with increased training and resistance to extinction. Additionally it appears to offer a basis for interpreting a recent negative finding (Wike, Kintsch, & Gutekunst, 1959).

From what has been said, it would appear that training and resistance to extinction are related in a curvilinear manner (Stimmel & North, 1959). The initial tendency toward increased resistance to extinction can be understood on the basis of long accepted learning theoretic principles, e.g., Hull (1943). The decreasing portion of the curve appears to require the assumption of a mediating reaction (Reid, 1953).

Regarding the use of nonlearners in statistical comparisons, several considerations may be raised which suggest that the conclusions drawn as a result of such comparisons do not find their basis in uncontrolled and unknown variables associated with selection. The available literature dealing with speed of learning and resistance to extinction

as reviewed by Hilgard and Marquis (1940), and Brogden (1951) strongly indicates that the correlation between the two variables is either zero or negative but not positive. This knowledge originally led us to the conclusion that learner-nonlearner comparisons were scientifically feasible. The experiment itself has since provided additional evidence consistent with the above view. It may be summarized as follows. Among learners, a correlation of −.22 obtained between speed of learning and resistance to extinction with trials overtraining partialed out. This nonsignificant correlation encourages the belief that the correlation in question was essentially zero over the entire range of *Ss*, learners and nonlearners alike. Further, the most highly trained learners were also in general the slower learners, yet these *Ss* extinguished most rapidly. Finally, fast learners extinguished rapidly only when overlearning trials were involved. Each of these facts independently suggests that a positive correlation between speed of learning and resistance to extinction had little or no effect on the present results and that overlearning was a potent variable in this investigation.

Summary

Four groups of 22 *Ss* each were trained using a regular pattern of events. Either 12, 24, 48, or 60 acquisition trials were given. Immediately following the last acquisition trial extinction training was begun. Increased training was found to decrease resistance to extinction but not significantly. The *Ss* who failed to learn the pattern of events were found to be most resistant to extinction. The *Ss* who learned the pattern proved to be least resistant to extinction provided they had received overlearning trials. Learning per se did not seem to reduce resistance to extinction. It was suggested that increased training results in reduced resistance to extinction only if such training includes overlearning trials. It was also suggested that the relationship between training and extinction is curvilinear and that the decreasing portion of the curve may require the assumption of a mediating reaction such as the one proposed by Reid.

References

Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 568–612.

Capaldi, E. J. The effects of different amounts of alternating partial reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1957, **70,** 451–452.

Capaldi, E. J. The effects of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, **51,** 367–371.

Goodnow, J. J., & Pettigrew, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51,** 385–392.

Harlow, H. F. Primate learning. In C. P. Stone (Ed.), *Comparative psychology.* (3rd ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1951. Pp. 183–238.

Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Lewis, D. J., & Duncan, C. P. The effect of partial reinforcement and length of acquisition-series upon resistance to extinction of a motor and a verbal response. *Amer. J. Psychol.*, 1956, **69,** 644–646.

Lewis, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55,** 121–128.

Perin, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30,** 93–113.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46,** 107–112.

Stimmel, D. T., & North, A. J. Effects of extended acquisition training on subsequent extinction of a running habit. *Amer. Psychologist*, 1959, **14,** 420.

Wert, J. E., Neidt, C. O., & Ahmann, J. S. *Statistical methods in educational and psychological research.* New York: Appleton-Century-Crofts, 1954.

Wike, E. L., Kintsch, W., & Gutekunst, R. Patterning effects in partially delayed reinforcement. *J. comp. physiol. Psychol.*, 1959, **52,** 411–414.

Williams, S. B. Resistance to extinction as a function of number of reinforcements. *J. exp. Psychol.*, 1938, **23,** 506–522.

(Received March 4, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 350-351

SUPPLEMENTARY REPORT: RESISTANCE TO EXTINCTION OF A VERBAL RESPONSE AS A FUNCTION OF THE NUMBER OF ACQUISITION TRIALS

MONTE G. SENKO, RONALD A. CHAMP, AND E. J. CAPALDI

University of Texas

Murillo and Capaldi (1961) found that under certain conditions increased training resulted in reduced resistance to extinction. The earlier studies of Williams (1938) and of Perin (1942) demonstrated increased resistance with increased training. Taken together the above results suggest that a ∩-shaped function describes the relationship between amount of training and resistance to extinction. In this study, training values were selected so as to make likely the appearance of the ∩-shaped function within the context of a single experimental setting.

Method.—The same apparatus and instructions were used as in the Murillo and Capaldi (1961) study. Eighteen *Ss* were assigned to each of eight groups, making a total of 144 *Ss*. Each group received a fixed, but different, number of 100% reinforced acquisition trials immediately followed, in each case, by 20 trials of extinction training. An acquisition trial was defined as one in which a piece of cloth was in the well of the apparatus; an extinction trial was defined by the absence of the cloth in the well. The numbers of acquisition trials given were: 0, 1, 3, 6, 12, 20, 40, or 80 trials. A guess of "in" is termed a positive response.

Results and discussion.—The Trial 1 percentage in the composite acquisition curve shown in Fig. 1 is based on data obtained from 144 *Ss*, Trial 2 on 108 *Ss*, and so on. The *Ss* responded at about the 100% level by Trial 10 and continued to respond at about that high level on subsequent trials.

In Fig. 2 are presented the data on the percentage of positive responses given during extinction by each of the groups except the 40- and 80-trial groups. The highly trained groups extinguished in a manner almost identical to that of the 20-trial group. Only

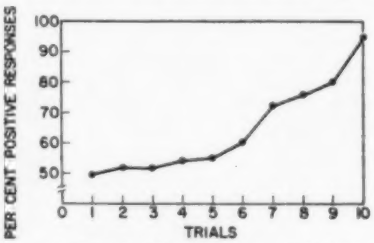

FIG. 1. Percent positive responses on each of the first 10 trials of acquisition.

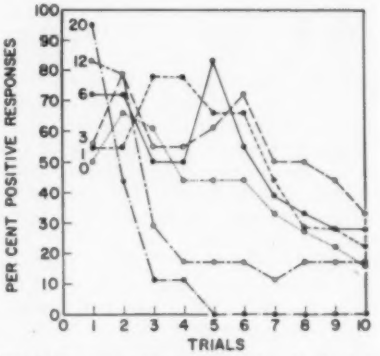

FIG. 2. Percent positive responses given by six of the eight groups on each of the first 10 trials of extinction.

the initial 10 trials are shown because, with the exception of the 3- and 6-trial groups, each group responded at practically a zero level beyond Trial 10. On Trials 11 through 16 the 3-trial group averaged about 30% positive responses and on the remaining trials about 15%. The 6-trial group maintained a fairly steady output of about 15% positive responses on the remaining 10 trials. As Fig. 2 shows, each group which received less than 12 acquisition trials tended to give an increased number of positive responses somewhere later in the extinction trials. For example, on Trials 3 and 4 of extinction, the 6-trial group responded at about the 50% level, but on Trial 5, the incidence of positive responses increased to about 83%. Some orderliness is apparent inasmuch as, with the exception of the 3-trial group, the increases appeared later in the extinction trials the more highly trained the group. A comparison of Fig. 1 and Fig. 2 will indicate that in some cases groups undergoing extinction responded at a higher level than they had in acquisition on comparable trials. For example, the 1-trial group gave approximately 78% positive responses on Trials 3 and 4 of extinction, while the incidence of positive responses on the comparable acquisition trials (Trials 4 and 5) was only about 55 and 60% respectively. Similar occurrences can be noted in the case of the 0- and 3-trial groups.

In Table 1 are presented the data on the mean number of positive responses given by each group in extinction. The number of

TABLE 1

MEAN NUMBER OF POSITIVE RESPONSES GIVEN BY EACH GROUP IN EXTINCTION

Measure	Group							
	0	1	3	6	12	20	40	80
R+	4.77	6.33	8.11	6.77	3.66	1.72	2.38	1.88
SD	2.55	2.36	2.13	2.54	1.26	0.79	1.38	1.07

positive responses steadily increased from the 0- to the 6-trial group whereupon a decreasing trend became manifest, reaching a minimum value at 20 trials and apparently stabilizing at the minimum value beyond 20 trials. A $\sqrt{X + .5}$ transformation was used. Analysis of variance indicated that the training variable was significant beyond the .001 level ($F = 32.36$, $df = 7/136$). Orthogonal subgroup comparisons indicated that resistance increased significantly from the 0- to the 1-trial group ($F = 6.17$, $df = 1/136$, $P < .05$), decreased significantly from 6 trials to 12 trials ($F = 22.20$, $df = 1/136$, $P < .05$), and that the difference between the 20- and 40-trial groups was unreliable ($F = 2.01$).

The results support the hypothesis that the relationship between amount of training and resistance to extinction is described by a ∩-shaped function. The rather high degree of resistance shown by the limited-training groups relative to the highly-trained groups is probably due in some part to the pre-experimental instructions. The present results are both in agreement and in disagreement with the results of Murillo and Capaldi. Agreement centers about the fact that increased training resulted in reduced resistance to extinction. However, in the Murillo and Capaldi study reduced resistance occurred only after the event pattern was well learned, whereas in the present study it began to occur with the 6-trial group, prior to the time that 100% positive responses occurred. The most plausible explanation for this discrepancy would appear to involve factors concerned with differences in the type of event pattern employed in the two studies. In the Murillo and Capaldi study the cloth was in the well on every third trial, out on the preceding two trials, whereas in this study a continuous pattern was employed.

REFERENCES

MURILLO, N. R., & CAPALDI, E. J. The role of over-learning trials in determining resistance to extinction. *J. exp. Psychol.*, 1961, **61**, 336–340.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

WILLIAMS, S. B. Resistance to extinction as a function of number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–522.

(Received March 28, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 4, 351–352

SUPPLEMENTARY REPORT: SPONTANEOUS RECOVERY OF PROBLEM SOLVING SET

WILLARD N. RUNQUIST AND BERT SEXTON

Pomona College

A previous study (Gardner & Runquist, 1958) produced evidence for the operation of the laws of simple associative learning and extinction in problem solving set. In the present study, an attempt was made also to demonstrate spontaneous recovery of a set which has been extinguished.

Method.—The problems used were the Water Jar Problems devised by Luchins (1942). Thirteen problems were used, 12 having the same method of solution (Solution A) and one having a different method of solution (Solution B). Four of eight groups were extinction groups, and received 11 Solution A problems followed immediately by the Solution B (extinction) problem. They received the final Solution A (test) problem after either 0, 10, or 30 min., or 48 hr. The remaining four groups served as controls for forgetting and were treated exactly the same as the experimental groups except that they received no extinction problem.

The *Ss*, 160 Junior College students randomly assigned 20 to a condition, were tested in groups of 30 to 50 in their regular classroom. Problems were presented in 5 × 7-in. booklets with one problem on each page. Each *S* kept his own solution time by looking at an assistant who indicated the elapsed time in 5-sec. intervals on the blackboard in front of the room. Detailed instructions including three examples were read to *Ss*. In addition, they were told to work each problem successively until they came to the part marked "rest period." No specific rest interval activity was used, as each *S* was told he could do anything during the rest period except look at or work previous problems.

Results.—The *Ss* recorded their solution times to the nearest 5 sec. These data

TABLE 1

MEAN LOG SOLUTION TIME ON TRAINING, EXTINCTION, AND TEST PROBLEMS AS A FUNCTION OF TIME INTERVAL

Group	Prob. 11	Prob. 12	Prob. 13	σ_m	Adj. Means Prob. 13
Experimental					
0-min.	1.43	1.82	1.61	.061	1.62
10-min.	1.44	1.92	1.47	.062	1.47
30-min.	1.47	2.00	1.53	.036	1.50
48-hr.	1.45	1.90	1.46	.055	1.44
All	1.45[a]	1.91[b]	1.51		1.51
Control					
0-min.	1.46		1.43	.042	1.40
10-min.	1.51		1.41	.036	1.33
30-min.	1.40		1.42	.035	1.46
48-hr.	1.32		1.44	.048	1.56
All	1.43[c]		1.43		1.44
Total	1.44		1.46		1.46

Note.—Hartley's test for homogeneity of variance on Prob. 13 did not produce evidence of heterogeneity ($F_{max} = 3.32$, $P > .05$).
[a] estimated $\sigma_m = .044$.
[b] estimated $\sigma_m = .087$.
[c] estimated $\sigma_m = .030$.

were transformed to log time in order to keep them comparable to those in the previous study. The mean log solution times for the last training problem, the extinction problem, and the spontaneous recovery test problem are given in Table 1. Despite the use of a group testing procedure and a different population of *Ss*, the means on the last training problem and on the extinction problem are very close to those obtained by Gardner and Runquist (1958). Inspection of the means in Problem 11 suggests that the groups may differ widely in performance on this problem. This was corroborated in an analysis of variance performed on these data ($F = 3.49$, $P < .01$). In order to correct for these initial differences, the means on the test problem were corrected by covariance. These adjusted means are also presented in Table 1.

The predictions concerning spontaneous recovery generally are upheld. Although there is some evidence of forgetting in the control groups, the groups which had the extinction problem actually improved with the passage of time. These conclusions are supported by an analysis of covariance which produced significant F for interaction between Time Interval and Extinction vs. Control ($F = 8.04$, $df = 3/151$). The main effects of Extinction vs. Control and Time Interval were also significant, but the latter being so only at the .05 level (F for Extinction is 7.73, $df = 1/151$; F for Time Interval is 4.12, $df = 3/151$).

Some specific comparisons between means were made with a t test, using an error term derived from the covariance analyses by means of a formula given by Cochran and Cox (1957, p. 78). These analyses showed that the differences favoring the control groups after 0 and 10 min. were significant ($t = 4.23$ and 2.69, respectively), while the differences after 30 min. was not significant ($t = .76$). The extinction group actually performed better than its corresponding control group after 48 hr. ($t = 2.31$). There is no readily apparent explanation for this phenomenon. Comparison of the 0-min. and 48-hr. groups within each condition showed a significant improvement in performance for the extinction groups and significant deterioration in the controls ($t = 3.46$ and 3.07, respectively).

In a descriptive sense, then, spontaneous recovery has been clearly demonstrated. It remains for further experimental and theoretical analysis to determine the extent to which the results of this experiment are due to the same factors which produce spontaneous recovery in simpler associative learning situations.

REFERENCES

COCHRAN, W. L., & COX, G. *Experimental design.* New York: Wiley, 1957.
GARDNER, R. A., & RUNQUIST, W. N. Acquisition and extinction of problem solving set. *J. exp. Psychol.*, 1958, **55**, 274–277.
LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**(6, Whole No. 248).

(Received September 26, 1960)

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland
PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich
PROF. DR. H. ROHRACHER, Wien

Herausgeber für die Schweiz
PROF. DR. R. MEILI, Bern

Redaktion
DR. C. J. HOGREFE, Göttingen

Band VII/4

WINFRIED LOHR, **Verhalten und Spannungszustände bei Säuglingen im Alter von zehn bis achtzehn Wochen in einer Wahrnehmungssituation**

THEO HERRMANN, **Über Urteilskonkordanz und Urteilsnuanciertheit**

MARTIN IRLE, **Eine Analyse von Beziehungen zwischen verwandten Einstellungen und Kenntnissen über den Gegenstand der Einstellungen**

HORST SACHER, **Regulierungspsychologische Persönlichkeitsdiagnostik**

ANNELIESE MERTN, **Validitätsuntersuchung zum soziometrischen Test**

HANS SZEWCZYK, **Zur Frage des Strukturwandels der Persönlichkeit in der Psychopathologie**

VICTORIA BRANDNER, **Dynamik und Ineinanderwirken von Einstellungen**

CHRISTIAN LOESCH UND FRIEDRICH SIXTL, **Ein elektronisches Gerät zur Registrierung von Sprechreaktionen**

DIETER V. KLEBELSBERG, **Diskussionsbemerkung zu zwei Abhandlungen über das Kieler Determinations-Gerät**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.
Subscription is DM 64.—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
Brentanoweg 10, Germany

Wiley BOOKS

Coming in May . . .

MOTIVATION AND EMOTION

A SURVEY OF THE DETERMINANTS OF HUMAN AND ANIMAL ACTIVITY

By Paul Thomas Young, *Professor of Psychology, Emeritus, University of Illinois.*

This is the only comprehensive textbook in the combined fields of motivation and emotion. While there is a current tendency to treat the topic of emotion within the context of motivation, or to treat emotion and motivation in closely related chapters, the present work recognizes the close relationship of both aspects and logically treats the two in a single volume. A *multifactor* approach to the analysis of the determinants of human and animal activity is employed which provides a broad scope and tolerates rather widely divergent views.

Motivation and Emotion contains the most recent and authentic account of the hedonic theory of motivation. The exposition of the author's research is focused toward the relationship with other workers in the field and clearly relates affectivity to other aspects of motivation.

Special features . . .

- Coverage is broader and more comprehensive than the coverage of other books in the field
- Integrates the topics of motivation in a new way
- Provides an authentic presentation of the author's hedonic theory of motivation and of the research leading up to this theory
- Presents the writer's views on psychological relativism

1961. **Approx. 640 pages.** **Prob. $10.50.***

**Textbook edition also available for college adoption*

JOHN WILEY & SONS, Inc.

440 Park Avenue South, New York 16, N. Y.